The coevolutionary dynamics between bacteria with CRISPR-Cas immunity and their lytic phages

Submitted by

Daniel Morley

to the University of Exeter as a thesis for the degree of

Masters by Research in Biological Sciences

September 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Abstract

Bacteriophage (phage), viruses of bacteria, are the most abundant and diverse biological entities on Earth. There are an estimated 10^{30} phages in the biosphere; collectively they outnumber their prokaryotic hosts tenfold, and they are thought to destroy up to half of the World's bacteria every 48 hours. Yet, despite the ubiquity of phage and the influence they exert over bacterial populations, we lack a detailed understanding of the complex bacteria-phage interactions that govern the coexistence of both species. Lytic phages – which require cell death to complete their life cycle – impose continuous selection pressures on bacteria to evolve resistance or face extinction. Likewise, increasing bacterial resistance creates strong selection pressures on phage to evolve greater infectivity. These reciprocal adaptations can lead to coevolution, which has profound consequences for the genetic diversity and evolutionary trajectories of both species.

One such phage-resistance mechanism in bacteria is CRISPR-Cas (clustered regularly interspaced short palindromic repeats - and associated Cas proteins); a heritable, adaptive immune system that is found in approximately half of all bacteria. CRISPR provides immunity against phages by incorporating phage-derived 'spacer' sequences into CRISPR loci on the bacterial genome. Transcriptions of these spacers can recognise and cleave complementary DNA from invading phage genotypes, but phage can 'escape' CRISPR by mutating the regions targeted by the spacers, enabling them to re-infect previously-resistant bacteria and potentially leading to coevolution. To date, very few CRISPR-phage systems have been studied in detail, but two of the best-known systems; *Pseudomonas aeruginosa* and *Streptococcus thermophilus*, show great variation in their efficacy at preventing phage infections. Whilst phage are rapidly driven extinct in *P. aeruginosa*, they can coexist over long periods in *S. thermophilus*; but it is unclear whether this phage persistence is due to coevolution, and if so, what kind of coevolutionary dynamics are associated with this system.

In this thesis, I examine *P. aeruginosa* and *S. thermophilus* together with their lytic phages in a coevolutionary context to better understand the role of

coevolution in CRISPR-phage interactions. I first explain how the diversity of CRISPR spacers generated in *P. aeruginosa* rapidly drives phage extinct, but that phage can become locally adapted when CRISPR spacer diversity is low. I then present the first empirical evidence of CRISPR-phage coevolution in *S. thermophilus* and show that coevolution follows an arms race dynamic in this system. I conclude by discussing the importance of CRISPR spacer diversity in determining the outcome of CRISPR-phage interactions and how low spacer diversity can lead to coevolution and local adaptation.

Contents

Tables and Figures

Chapter 1: General Introduction

Chapter 2: Host diversity limits the evolution of parasite local adaptation.

Chapter 3: CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage.

Table 2 The mean proportion of hosts infected and phage resisted in pairwise challenges in the time-shift experiment, broken down by the day from which the host or phage originated.

Table 3 Contingency table of pairwise infections that were predicted to lead to phage escape based on protospacer sequence data, compared against the pairwise infections that were measured from the phenotypic assay.

S. Figure 1 The proportion of hosts resistant to phage that were from the host's past, present or future.

S. Figure 2 The relative importance of fluctuating selection dynamics to arms race dynamics in this model system.

S. Figure 3 The relationship of host susceptibility to (a) the number of spacers acquired per clone, and (b) the sequence diversity in terms of the pairwise difference among spacer sequences from the phenotypic data.

S. Table 1 Infectivity matrices from the phenotypic assay.

S. Table 2 Unique host CRISPR spacers detected by PCR analysis.

S. Table 3 Locations of SNPs detected by PCR when mapped to the phage 2972 genome.

Foreword

Some of the research published in this thesis was a collaboration between multiple researchers. The individual researcher contributions were as follows:

Chapter 2: Host diversity limits the evolution of parasite local adaptation.

Authors	**Daniel Morley (DM),** Jenny M. Broniewski (JB), Edze R. Westra (EW), Angus Buckling (AB), Stineke van Houte (SvH)
Contributions	SvH, EW, AB and **DM** designed the experiments. **DM**, JB and SvH performed the experiments. **DM** and SvH analysed the data. All authors contributed to the manuscript.
Reference	Morley, D., et al. (2017) "Host diversity limits the evolution of parasite local adaptation." *Molecular ecology* 26.7: 1756-1763.

Chapter 3: CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage

Authors	Jack Common (JC)†, **Daniel Morley (DM)**†, Edze R. Westra (EW), Stineke van Houte (SvH)
Contributions	SvH and EW designed the experiments. **DM**, JC, SvH and EW performed the experiments. JC analysed the data. All authors contributed to the manuscript. †Contributed equally.
Reference	Common, J., et al. (2018) “CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage” *Manuscript submitted.*

Abbreviations

AIC	Akaike information criterion. An estimation of the relative quality of statistical models for a given set of data.
ARD	Arms race dynamic. A type of antagonistic coevolution characterised by increasing levels of host resistance and parasite infectivity.
cfu	Colony-forming units per millilitre. A measurement of bacterial density; usually expressed as the number of colony-forming units per millilitre (cfu ml^{-1}).
CRISPR-Cas	Clustered regularly interspaced short palindromic repeats - and associated Cas proteins. An adaptive prokaryotic immune system.
CRISPR-KO	CRISPR-knockout. A strain of bacteria that carries a non-functional CRISPR-Cas system.
crRNA	CRISPR RNA. RNA transcribed from the CRISPR array that combines with Cas proteins to form a ribonucleoprotein complex involved in phage interference.
dpi	Days post-infection.
FSD	Fluctuating selection dynamic. A type of antagonistic coevolution characterised by selective sweeps of rare genotypes.
G×E	Genotype-Environment Interaction. A framework used to model the effects of genotype and environment on an organism.
GFG	Gene-for-gene. A type of infection genetics underpinning host-parasite coevolution whereby hosts and parasites adopt a broad resistance / infectivity range.

GLM General linear model. A statistical model for comparing how several variables affect different continuous variables.

GLMM General linear mixed model. An extension of a GLM which allows for the additional modelling of random effects.

M17 M17 media. Bacterial growth medium made from M17 broth supplemented with α-Lactose.

M9 M9 media. Bacterial growth medium made from M9 salts supplemented with glucose.

MA Matching alleles. A type of infection genetics underpinning host-parasite coevolution whereby hosts and parasites specialise in resisting / infecting a single genotype.

OD_{600} Optical density. A measure of the amount of light (at 600 nm) that is refracted from a sample. Used to approximate bacterial densities.

PAM Protospacer adjacent motif. A short, highly-conserved sequence of nucleotides on the phage genome that is used by bacteria to distinguish self from non-self DNA during CRISPR spacer acquisition.

PCR Polymerase chain reaction. A technique used to amplify sections of DNA.

pfu Plaque-forming units. A measurement of phage titre; usually expressed as the number of plaque-forming units per millilitre (pfu ml^{-1}).

PWD Pairwise difference. A measure of the difference between two or more pairs of related values.

rpm Revolutions per minute.

sm Surface mutant. Bacteria that have modified their surface proteins to prevent phage adsorption.

SNP Single nucleotide polymorphism. A mutation in a single nucleotide along a stretch of DNA.

spp. Species pluralis. Multiple species belonging to the same genus.

WT Wild-type. The phenotype of a typical strain of bacteria as it occurs in nature.

Acknowledgements

I would like to thank my family for their constant encouragement and unwavering support during the course of my studies; especially Emma, who always had time to listen, and Peter Rudin, Derek Howe and my dad Graham for keeping Viridian Media running smoothly in my absence and enabling me to disappear from work at short notice for extended periods of time.

I would also like to thank Professor Angus Buckling for giving me the opportunity to conduct this research despite my unorthodox route into microbiology, and my friends and colleagues at the ESI for making the laboratory a pleasurable and stimulating place to work.

Finally, this thesis would not have been possible without the expert knowledge and supervision of Dr Stineke van Houte and Dr Edze Westra. Thank you both for your patient guidance over these past two years, and for giving me confidence in my abilities as a researcher. *Gaan met die banaan!*

Chapter 1

General Introduction

Coevolution is a fundamental tenet of biological science and may be as important as local natural selection in determining the evolutionary trajectories of species due to the influence it exerts on gene transfer and population dynamics. Coevolution occurs when two or more species evolve reciprocal adaptations in response to one another. There are countless examples of coevolution in biological literature; such a broad topic naturally covers a diverse range of species and ecologies, and often the evolution of adaptive and counter-adaptive traits can lead to complex coevolutionary dynamics between species that result in permanent interdependent relationships.

Occasionally, coevolution can benefit both species simultaneously; for example, the Bullhorn Acacia tree (*Acacia cornigera*) has evolved a complex symbiotic relationship with the ant species *Pseudomyrmex ferruginea*; the Acacia provides ants with domicile and food in exchange for protection against phytophagous insects and neighboring plants (Janzen, 1966), however; coevolutionary adaptations are more commonly antagonistic and evolve in response to interactions between competing, predatory, parasitic, or otherwise-opposed species. In a microbiological context, antagonistic host-parasite coevolution is frequently observed between bacteria and bacteriophage (phage) viruses as they compete for survival. Phages are abundant in nature, and their large population sizes, short generation times, and fast migration rates (relative to bacteria) make them a persistent hazard to bacterial communities. This is especially true for lytic phages which are obligate killers that destroy any cells they infect (Figure 1). To cope with the constant threat of infection, bacteria have evolved a multitude of phage defence mechanisms that can be broadly classified as either constitutive (i.e. permanently active) or inducible (i.e. activated upon phage exposure) (Tollrian *et al.* 1999). Both strategies have differing fitness costs and consequently their evolution is driven largely by nutrient availability and exposure to phage, with constitutive defences evolving under high-nutrient, high-phage conditions, and vice versa for inducible defences (Westra *et al.* 2015). Common examples of constitutive defence mechanisms include surface modification, for

example by modifying or down-regulating cell-surface receptors to prevent phage adsorption (Labrie *et al.* 2010; Westra *et al.* 2012), and restriction-modification where unmethylated phage DNA is degraded using endonuclease enzymes (Wilson & Murray 1991). Inducible defences, which are triggered by prior exposure to phage, include preventing additional phage DNA from entering the cell (superinfection exclusion) (Labrie *et al.* 2010), inducing cell suicide to limit the spread of infection (abortive infection) (Fineran *et al.* 2009), and cleaving complementary phage DNA based on 'memory' of prior infections (CRISPR-Cas) (Westra *et al.* 2012).

CRISPR is the most recently-discovered of these responses and is unique in immunological terms because it represents the first known example of heritable, adaptive immunity in bacteria (Sorek *et al.* 2008). CRISPR is a useful tool for studying bacteria-phage interactions because it rapidly evolves resistance when exposed to phage (Paez-Espino *et al.* 2015) which means coevolutionary interactions can be measured in 'real-time' as opposed to alternative methods that infer coevolution based on past records of resistance and infectivity genes. In addition, the evolution of CRISPR resistance occurs in a predictive way at known locations on the host's genome (Datsenko *et al.* 2012) which means it is easy to quantify and conducive to mathematical models that can be used to simulate coevolutionary dynamics. CRISPR systems are also ubiquitous in natural bacterial populations and can theoretically be used to study coevolutionary interactions in bacteria-phage systems that cannot be cultured in a laboratory.

An overview of CRISPR-Cas

CRISPR was first described in 1987 by Ishino and colleagues at Osaka University who accidentally cloned part of a CRISPR array whilst working on an unrelated gene in *Escherichia coli*. The sequence they identified; five short homologous repeats interspaced by equally-short heterologous sequences (Ishino *et al.* 1987), had no known function at the time, and its biological significance remained unclear until 2002 when Jansen *et al.* (2002) found similar sequences in other prokaryotic genomes and discovered that they co-localised with adjacent *cas* genes. Three years later, Mojica *et al.* (2005) discovered that the interspacing heterologous sequences were in fact derived from foreign genetic elements,

which eventually led to the discovery that CRISPR functions as an immune response in prokaryotes (Barrangou *et al.* 2007).

CRISPR systems are widespread in nature and can be found in approximately half of all bacteria (Grissa *et al.* 2007). They are also highly diverse, and are currently ordered into two classes, six types, and 33 subtypes (Chaudhry *et al.* 2018; Tollrian *et al.* 1999; Vale *et al.* 2015) depending on their genetic composition and repeat sequences. CRISPR loci consist of an array of phage-derived 'spacer' sequences of around 30 bp which are interspaced by homologous repeat sequences of a similar length. The array is flanked by a 'leader' sequence that contains promoter elements and regulatory binding sites for associated *cas* genes that lie further upstream (Childs *et al.* 2012; Labrie *et al.* 2010; Westra *et al.* 2012; Wilson & Murray 1991) (Figure 2a). All CRISPR systems share a similar mechanism of action that can be described in three distinct stages: adaptation, expression, and interference (van Houte *et al.* 2016a) (Figure 2b). During adaptation, Cas proteins form a complex that binds to the target sequence on the phage genome (the 'protospacer') and inserts it into the host's CRISPR array as a spacer (Labrie *et al.* 2010; Westra *et al.* 2012). During expression, Cas proteins are up-regulated, and spacer sequences from the CRISPR array are transcribed into RNA (crRNA) containing the complementary spacer sequence derived from the phage protospacer (Van der Oost *et al.* 2009). During interference, crRNAs bind with other Cas proteins to form a CRISPR-Cas complex that can recognise and cleave complementary DNA from invading phage (Marraffini & Sontheimer 2010). The acquisition of CRISPR spacers during adaptation is a non-random event that depends upon the presence of a protospacer adjacent motif (PAM); a highly-conserved, 2-7 nt sequence at the 3' end of the protospacer on the phage genome (Deveau *et al.* 2008). The PAM acts as a binding site to help guide Cas proteins during spacer acquisition and is important for bacteria to distinguish between self and non-self DNA (i.e. to prevent self-cleavage).

Since CRISPR relies on complementary base pairing between the crRNA-Cas complex and the invading phage DNA to prevent infection, phage can overcome or 'escape' CRISPR by evolving single point mutations (SNPs) in the PAM sequence or protospacer region targeted by the spacers (Deveau *et al.* 2008)

which enables them to re-infect previously-resistant hosts. In return, bacteria can acquire multiple CRISPR spacers (van Houte *et al.* 2016b), each targeting a different phage genotype (or a different part of the same phage genotype), potentially resulting in a coevolutionary arms race with increasing CRISPR spacer acquisition and subsequent phage point mutations over time

CRISPR-Cas infection genetics and coevolutionary dynamics

Where hosts and parasites coevolve together, their pattern of coevolution typically follows either an arms race dynamic (ARD) which is characterised by escalating levels of host resistance and parasite infectivity, or a fluctuating selection dynamic (FSD) which is based on negative frequency-dependent selection resulting in 'selective sweeps' of rare genotypes (Gandon *et al.* 2008) (Figure 3). The genetic basis behind these two systems is often modelled as a spectrum with a gene-for-gene (GFG) model at one end and a matching alleles (MA) model at the other (Agrawal & Lively 2002). Under GFG, hosts and parasites evolve as 'generalists' that can resist / infect a broad range of genotypes, resulting in escalating resistance and infectivity over time (ARD). In contrast, hosts and parasites in MA systems evolve to resist / infect a single genotype (Agrawal & Lively 2002) (Figure 4) which leads to FSD as rare genotypes can proliferate and spread throughout the population. Furthermore, MA can give rise to local adaptation where hosts 'specialise' in resisting parasites (and parasites specialise in infecting hosts) from the same geographic and temporal environment (Gandon & Nuismer 2009; Lively 1999; Morgan *et al.* 2005).

CRISPR conforms to a MA model of infection genetics due to the high degree of specificity required between the CRISPR spacer and the invading phage genome. Based on common evolutionary theory (above), this should lead to FSD as bacteria acquire spacers that provide resistance against the most common phage genotype(s), and in return phage 'escape' CRISPR by mutating the regions targeted by the spacer (Tyson & Banfield 2008; Westra *et al.* 2016). However, it is possible for bacteria to acquire multiple CRISPR spacers (and phage to evolve multiple point mutations) which could lead to the evolution of more generalist resistance / infectivity genotypes and could potentially result in ARD (Vale & Little 2010).

Modelling of CRISPR-phage interactions suggests that CRISPR can lead to stable coexistence (Iranzo *et al.* 2013), however the phenomenon is difficult to study empirically due to a lack of suitable model systems. In natural environments, phages have been found to coevolve with *Leptospirillum* spp. in biofilms sampled from acidic wastewater at a sulphide mine in Northern California (Andersson & Banfield 2008), and in the fish pathogen *Flavobacterium columnare* at a fish farm in Central Finland (Laanto *et al.* 2017) which suggests that CRISPR-phage coevolution might be commonplace in nature. However, coevolution has never been directly observed under laboratory conditions, and previous studies have yielded mixed results. For example, in *Pseudomonas aeruginosa*, phages are unable to coevolve and are rapidly driven extinct by CRISPR; an effect that is attributed to the high levels of spacer diversity generated in the host population (van Houte *et al.* 2016b). In contrast, phages have been shown to coexist for long periods with *Streptococcus thermophilus* (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015), although it is unclear whether this coexistence was due to coevolution, or whether it could be explained by some other mechanism such as loss of CRISPR immunity which would provide a continuous reservoir of sensitive hosts for the phage to amplify in (Chaudhry *et al.* 2018; Weissman *et al.* 2018).

In this thesis, I examine the antagonistic CRISPR-phage interactions in two model systems; *Pseudomonas aeruginosa* and *Streptococcus thermophilus*, to better understand the role that CRISPR plays in a coevolutionary context. I first explain how coevolution does not occur between *P. aeruginosa* UCBPP-PA14 and phage DMS3vir due to the diversity of CRISPR spacers generated in the host population, but show that phages can become locally adapted when CRISPR spacer diversity is low (Morley *et al.* 2017). In the following chapter, I present the first empirical evidence of CRISPR-phage coevolution between *S. thermophilus* DGCC7710 and phage 2972 and show that coevolution in this model system follows an arms race dynamic (ARD). In the final chapter I discuss the generation, maintenance, and limitations of CRISPR spacer diversity, and how these factors can explain why coevolution occurs in one system but not the other. I conclude by offering thoughts on future research directions.

Figures

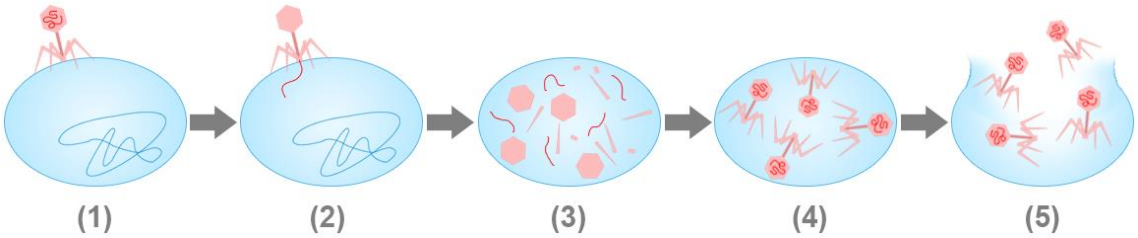


Figure 1. The life cycle of a lytic phage. (1) Phage adsorption to the surface of the cell. (2) Phage DNA injection. (3) DNA replication and biosynthesis of phage proteins (4) Phage protein assembly. (5) Cell lysis and release of new phage progeny.

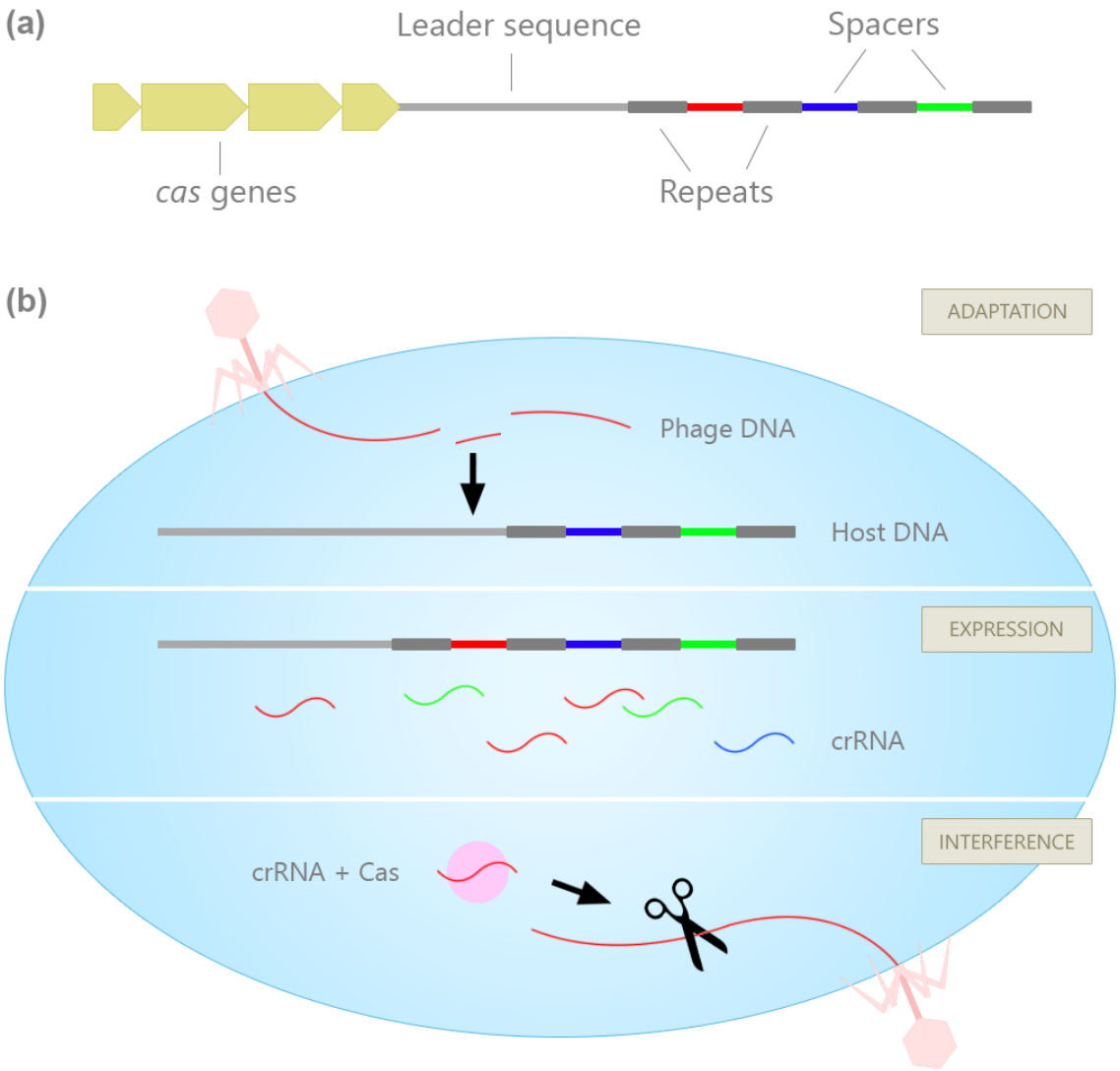


Figure 2. (a) Schematic of a CRISPR locus showing the position of the *cas* genes, leader sequence and CRISPR array. Multiple spacers can be acquired from different phage genomes (indicated here by different colours). Adapted from Labrie *et al.* (2010). **(b)** An overview of CRISPR-mediated immunity in bacteria. Phage-derived spacer sequences are inserted into the host's genome during adaptation. These spacer sequences are transcribed CRISPR RNAs

(crRNA) during expression and combine with Cas proteins to form a complex that can recognise and cleave complementary phage DNA during interference. Adapted from Westra *et al.* (2012).

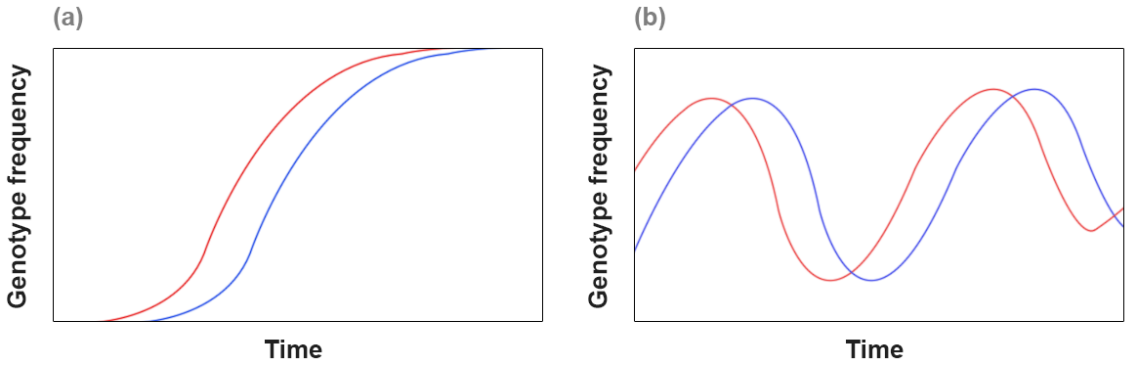


Figure 3. The frequency of bacteria (red) and phage (blue) genotypes during different coevolutionary dynamics. **(a)** During an arms race dynamic (ARD), bacteria and phage evolve increasing levels of resistance / infectivity over time. **(b)** During fluctuating selection dynamics, bacteria and phage undergo negative frequency-depending selection which results in selective sweeps of their respective genotypes.

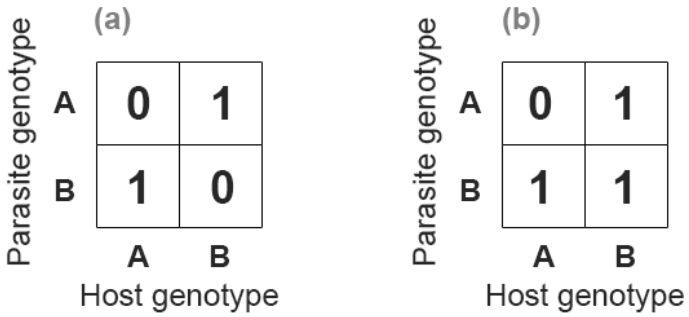


Figure 4. Infection and resistance success in **(a)** matching allele (MA) and **(b)** gene-for-gene (GFG) infection regimes. Under MA, hosts and parasites specialise to resist / infect a single genotype, i.e. parasite A can only infect host B and parasite B can only infect host A. Under GFG, hosts and parasites evolve general resistance and infectivity, i.e. parasite A can only infect parasite B but parasite B can infect hosts A and B. "1" indicates successful infection, "0" indicates successful resistance. Adapted from Råberg *et al.* (2014).

Chapter 2

Host diversity limits the evolution of parasite local adaptation

Abstract

Specificity in the interactions between hosts and their parasites can lead to local adaptation. However, the degree of local adaptation is predicted to depend upon the diversity of resistance alleles within the host population; increasing host diversity should decrease mean parasite infectivity and hence reduce local adaptation. In this study we empirically test this prediction using the highly specific interactions between bacteria with CRISPR-Cas immunity and their bacteriophage. Bacteria acquire immunity to phage by incorporating a phage-derived sequence into CRISPR loci on the host genome, and phage can escape the CRISPR-mediated immunity of a specific clone by mutating the targeted sequence. We found that high levels of CRISPR allele diversity which naturally evolve in host populations exposed to phage (because each bacterial clone captures a unique phage-derived sequence) prevents phage from becoming locally adapted. By manipulating the number of CRISPR alleles in the host population we show that phage can become locally adapted to their bacterial hosts but only when CRISPR allele diversity is low.

Introduction

Parasites are said to be locally adapted when they are more infective on hosts from the same geographic area (sympatric hosts) compared to hosts from different locations (allopatric hosts) (Kawecki & Ebert 2004). Variation in the sign and magnitude of parasite (and host) local adaptation across space has important consequences for coevolutionary dynamics (Thompson 1994) and the generation and maintenance of biodiversity (Thompson 1999), and local adaptation is likely to be a key determinant of invasion success (isolation by adaptation; Thompson (2005); Orsini *et al.* (2013)). In coevolving populations, the extent to which parasites are locally adapted is likely to depend upon the degree of host-parasite specificity (Lively 1999; Morgan *et al.* 2005). All else being equal, local adaptation is most likely to occur under conditions of very high specificity when parasites can infect a similar number of host genotypes in unique combinations, as opposed to when hosts and parasites vary in their infectivity and resistance ranges (Gandon & Nuismer 2009; Lively 1999; Morgan *et al.* 2005).

High infection specificity can result from interactions between host-parasite systems and the abiotic environment (Gandon & Nuismer 2009; Gorter *et al.* 2016; Laine 2008; Lopez Pascua *et al.* 2012), genetic constraints that limit parasite infection ranges (Agrawal & Lively 2002), or through fitness costs that inhibit the evolution of generalists (Gómez *et al.* 2015; Sasaki 2000). However, the importance of infection specificity on parasite local adaptation is thought to be affected by the diversity of infectivity / resistance alleles within each population. Assuming a parasite genotype needs to match the host genotype for infection to occur, increasing the number of host genotypes will reduce mean infectivity across all populations and hence the potential for local adaptation (Gandon & Nuismer 2009). The same argument holds if resistance is determined by matching; more parasite genotypes decreases mean resistance. Here, we investigate the importance of within-population diversity in the highly specific host-parasite interaction between CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-associated) and bacteriophages (phage) in determining local adaptation.

CRISPR-Cas provides hosts with an adaptive immune system that can target previously encountered phages (Barrangou *et al.* 2007). CRISPR loci consist of

an array of repeating sequences of approximately 30 bp interspaced by variable spacer sequences of similar length that are derived from foreign genetic material such as invading phage DNA. CRISPR-associated (*cas*) genes encode proteins that identify and degrade foreign DNA in a process which can be broadly defined in three stages: adaptation, expression and interference (van Houte *et al.* 2016a). During adaptation a new spacer sequence is acquired from the phage DNA and integrated into the CRISPR array (Amitai & Sorek 2016). At the expression stage the CRISPR array is transcribed into precursor CRISPR RNAs, which are subsequently cleaved into mature CRISPR RNAs (crRNA) each containing a single spacer sequence (Plagens *et al.* 2015). Lastly, during interference crRNAs guide Cas proteins to detect and cleave nucleic acids complementary to the spacer sequence (the protospacer) on the invading phage DNA (van der Oost *et al.* 2014). To discriminate between self and non-self DNA, most CRISPR-Cas systems rely on a highly-conserved sequence directly adjacent to the protospacer on the phage genome termed the protospacer-adjacent motif (PAM; see Deveau *et al.* (2008)). PAMs are absent from the host CRISPR loci, which prevents the host from cleaving its own genome. Subsequently, phage can escape CRISPR degradation by acquiring single point mutations in the protospacer or PAM (Semenova *et al.* 2011), however such mutations enable phage to overcome a single host genotype only and do not influence the infectivity on other host genotypes.

This high infection specificity can lead to ongoing coevolution (Agrawal & Lively 2002; Childs *et al.* 2014; Iranzo *et al.* 2013; Levin *et al.* 2013), which can theoretically result in the local adaptation of either bacteria or phage populations. Indeed, phage have been reported to persist in wild-type populations of *Streptococcus thermophilus* for at least 30 days (Paez-Espino *et al.* 2015), and several metagenomic studies have shown that CRISPR spacers tend to match sympatric phage genomes rather than phages from foreign environments (Berg Miller *et al.* 2012; Emerson *et al.* 2013; Held & Whitaker 2009; Kunin *et al.* 2008; Sorokin *et al.* 2010). However, experiments with *Pseudomonas aeruginosa* and the phage DMS3vir show that CRISPR-immune hosts rapidly drive phage extinct (van Houte *et al.* 2016b) due to the high levels of within-population diversity generated at CRISPR loci. This observation suggests that the diversity of

CRISPR alleles within the host population may limit the extent to which phage can become locally adapted.

In this study we used the CRISPR-Cas / phage model system to address the importance of host diversity in determining the extent of phage local adaptation. First, we carried out a natural coevolution experiment to assess the degree of local adaptation in WT populations. We anticipated that there would be little phage local adaptation due to the large amounts of host diversity generated by CRISPR-Cas. We then reanalysed data from previous short-term coevolution experiments in which we manipulated the levels of CRISPR spacer diversity to generate populations of between 1 and 48 unique host genotypes. We hypothesised that the high specificity of the CRISPR-Cas / phage interaction would result in strong local adaptation under conditions of low host diversity, but that phage local adaptation would decrease in the presence of more diverse host populations.

Methods

Bacterial strains and viruses

The *Pseudomonas aeruginosa* strains UCBPP-PA14 (WT), UCBPP-PA14 *csy3::LacZ* (CRISPR-KO) and the CRISPR-KO derived surface mutant (*sm*) were used throughout this study. The CRISPR-KO strain carries a non-functional CRISPR-Cas system, and the *sm* strain is a derivative which has acquired phage resistance by loss of the pilus (Westra *et al.* 2015). Phages DMS3vir (Cady *et al.* 2012) and DMS3vir+*acrF1* (Bondy-Denomy *et al.* 2013) are used as described in Westra *et al.* (2015) and van Houte *et al.* (2016b).

Coevolution experiments

Coevolution experiments were performed in twelve replicates by inoculating glass vials containing 6 ml M9 supplemented with 0.2% glucose (M9 media) with 10^7 colony-forming units (cfu) of WT *P. aeruginosa* from a fresh culture that had previously been adapted to M9 medium, and 10^4 plaque-forming units (pfu) of phage DMS3vir. Vials were then incubated at 37^{o}C while shaking at 180 rpm. Cultures were transferred daily 1:100 into fresh M9 medium for five days. Samples were taken every 24 hours just prior to each daily transfer and frozen in 20% glycerol at -80^{o}C. Phage extractions were performed by chloroform extraction and stored at 4^{o}C.

Evolution of immunity

To measure the type of bacterial immunity that had evolved during the coevolution experiments, 24 individual colonies were picked from each of the twelve replicates at 3 days post-infection (dpi) and streaked against ancestral phage DMS3vir and phage DMS3vir+*acrF1* which carries an anti-CRISPR gene that enables it to infect hosts which have evolved CRISPR immunity (Bondy-Denomy *et al.* 2013). Clones that experienced inhibited growth when exposed to DMS3vir+*acrF1*, but remained unaffected by ancestral DMS3vir, were deemed to have evolved CRISPR immunity, whereas clones that were resistant to both phages were deemed to have evolved surface modification (since CRISPR-immune hosts would still be sensitive to DMS3vir+*acrF1*). Clones susceptible to both phages were regarded as sensitive (i.e. no immunity had evolved).

Measuring phage local adaptation on coevolved populations

To determine phage local adaptation, cross-streak assays were performed in which the 24 individual bacterial clones from each replicate that were used to determine the type of bacterial immunity (see previous section) were streaked against each of the twelve phage populations extracted at 3 dpi. Where necessary, phage populations were amplified by infecting a liquid culture of sensitive CRISPR-KO bacteria overnight and harvesting amplified phage the following morning by chloroform extraction to obtain titres of at least 10^6 pfu ml^{-1}. Bacteria were regarded sensitive to a phage population if any inhibition of growth was observed, whereas clones that grew uninhibited were regarded as resistant. Local adaptation was analysed using the two methods described by Kawecki and Ebert (2004): 1) by subtracting the performance of allopatric (foreign) phages from the performance of the sympatric (local) phage for each clone (i.e. “foreign vs. local”) and 2) by subtracting the performance of a phage population on allopatric (away) hosts from the performance of that phage population on sympatric (home) hosts (i.e. “home vs. away”). Statistical analysis was performed in JMP software (v. 10.0.0).

Measuring phage local adaptation on host populations with varying CRISPR spacer diversity

Phage local adaptation was measured on host populations containing different levels of CRISPR spacer diversity as discussed in van Houte *et al.* (2016b). Briefly, we used 48 clones with unique CRISPR spacer sequences to generate populations that consisted of either a single CRISPR clone (monocultures; 48 replicates) or mixtures of 6 clones (8 different populations), 12 clones (4 different populations) or 24 clones (2 different populations) so that each population carried a unique complement of CRISPR spacers. Each of these populations was then combined 50:50 in competition assays with either a surface mutant (*sm*) which had acquired phage resistance through the loss of the pilus, or with a functional CRISPR-knockout population (CRISPR-KO) in which the phage could replicate. Phage samples were taken at regular time points and spotted onto every bacterial clone to monitor the emergence of escape phage over time (see van Houte *et al.* (2016b)). At 1, 2 and 3 dpi phage local adaptation was analysed by subtracting the performance of allopatric phages from the performance of the sympatric phage for each host (“foreign vs. local”), and by subtracting the performance of a phage population on allopatric hosts from the performance of that phage

population on sympatric hosts (“home vs. away”). The mean for each population was used as a measure of phage local adaptation; statistical analysis was performed in JMP software (v. 10.0.0).

Results

To understand how CRISPR-Cas is associated with local adaptation, we first performed a coevolution experiment using *P. aeruginosa* and its phage DMS3vir. Phage titres initially increased in all replicates during the first day after infection, after which phage titres decreased until they became extinct in all replicates by day five (Figure 1a). Bacterial densities increased from approximately 10^7 to 10^9 cfu ml^{-1} throughout the experiment. Analysis of the evolution of immunity showed that at 3 dpi the majority (77% ± 5% (mean ± 95% CI)) of bacteria had evolved CRISPR-mediated immunity, while 15% (± 6%) had evolved surface modification and 8% (± 2%) were still sensitive to the phage (Figure 1b). Analysis of cross-streak assays performed at 3 dpi revealed no significant phage local adaptation in any of the 12 replicates (Figure 2; t-test against zero: $t_{11} = 0.4135$, $p = 0.3436$), indicating that CRISPR-mediated resistance does not lead to phage local adaptation (or maladaptation) in this model system. We reasoned that this may be due to the high levels of spacer diversity which are commonly generated at the CRISPR loci of WT *P. aeruginosa* during CRISPR adaptation (van Houte *et al.* 2016b; Westra *et al.* 2015) as the phage are unable to overcome the diversity of spacer sequences in the host population.

We tested this hypothesis by studying local adaptation on host populations of varying CRISPR spacer diversity. We used 48 CRISPR clones with unique spacer sequences to generate the diverse populations described in the Materials and Methods. Each population was competed with a surface mutant (*sm*) or CRISPR-KO strain (van Houte *et al.* 2016b), and phage local adaptation was examined at 1, 2 and 3 dpi. Analysis of the *sm* competition experiment showed that escape phage (phage which had evolved point mutations in the sequence targeted by CRISPR, enabling them to overcome the CRISPR immunity of clones in the host population) appeared in 43 out of 48 monocultures, five out of eight 6-clone populations, and three out of eight 12-clone populations at 3 dpi. No escape phage were detected in the 24-clone population (see Figure 4 in van Houte *et al.* (2016b)). There was a clear inverse relationship between the levels of host diversity and the degree of phage local adaptation (Figure 3a; Linear Regression; $R^2 = 0.4066$, $F_{1,60} = 41.1122$, $p = <0.0001$). Phage quickly became locally adapted to monocultures (Figure 3a; t-test against zero: $T_{47} = 19.1301$, $p = <0.0001$) but as spacer diversity increased, the level of local adaptation decreased

dramatically. Local adaptation was still detected for the 6-clone population (t-test against zero: $T_7 = 2.2851$, $p = 0.0281$), but no local adaptation could be detected for the 12-clone or 24-clone populations. Consistent with these results, we also found that phage performance (as measured by mean infectivity) decreased with increasing spacer diversity (Figure 3b; Linear Regression: $R^2 = 0.090791$, $F_{1,122} = 12.1825$, $p = 0.0007$). Phage local adaptation was also measured for each population as the mean performance of each phage on its sympatric host minus its performance on allopatric hosts (i.e. “home vs. away”), however this alternative measure did not qualitatively change our results, which are summarised in Table 1.

To understand the effect of increased phage evolutionary potential on local adaptation, we also measured phage local adaptation on the same CRISPR-resistant host populations, but this time competed with a functional CRISPR knockout strain (CRISPR-KO) in which the phage could replicate (Supplementary Figure 1). Escape phage persisted in 45 out of 48 monocultures, eight out of eight 6-clone populations, and six out of eight 12-clone populations at 3 dpi, however no escape phage were detected in the 24-clone population (see Extended Data Figure 4 in van Houte *et al.* (2016b)). Local adaptation analysis showed comparable results to the *sm* competition experiment (experiment effect: $F_{1,120} = 0.6670$, $p = 0.4157$; experiment x diversity level effect: $F_{1,120} = 0.0141$, $p = 0.9055$): phage became strongly locally adapted to monocultures (t-test against zero: $T_{47} = 24.1288$, $p = <0.0001$) and to the 6-clone populations (t-test against zero: $T_7 = 5.2190$, $p = 0.0006$), but again no local adaptation was detected in the more diverse host populations regardless of which method was used for measurements.

Discussion

In this study we used the bacterium *P. aeruginosa* and its phage DMS3vir to test the hypothesis that under conditions of high infection specificity, parasite local adaptation decreases as the diversity of resistance alleles within the host population increases (Gandon & Nuismer 2009). First, we showed that coevolution does not lead to local adaptation in this model system due to the high levels of spacer diversity generated by CRISPR-Cas in response to phage. By manipulating the levels of host spacer diversity (van Houte *et al.* 2016b) we demonstrated that there is an inverse relationship between phage local adaptation and the diversity of CRISPR alleles within the host population. We attribute this relationship to the high within-population diversity of CRISPR spacers targeting the phage which increases overall mean resistance in the host population (Gandon & Nuismer 2009). Since phage mutations are required to escape CRISPR-Cas immunity, increasing spacer diversity naturally decreases the number of potentially sensitive hosts. Our data are entirely consistent with this mechanism; the mean infectivity levels of sympatric and allopatric phage populations decreased with increasing spacer diversity (Figure 3b). However, the lack of local adaptation at high levels of diversity may have arisen simply because the phage were driven extinct more rapidly, providing insufficient opportunities for the emergence of escape mutants and hence the detection of local adaptation. To disentangle these effects, we competed our host populations of varying spacer diversity with a sensitive CRISPR-KO strain in which the phage could replicate, and found qualitatively similar results (Supplementary Figure 1). Thus, local adaptation does not appear to be constrained by a lack of infective mutants; instead we conclude that local adaptation decreases as the diversity of resistance alleles increases due to the reduction in overall mean infectivity on CRISPR-immune hosts.

Parasite local adaptation is more frequently reported in studies where parasites migrate more than their hosts (Greischar & Koskella 2007). In our empirical system, it is likely that phage migration would lead to an increase in phage diversity, and this could increase local adaptation on CRISPR-resistant hosts since the host population would require a greater diversity of CRISPR spacers to overcome the phage. However, the lack of local adaptation in the synthetic CRISPR communities of 24 clones when phage evolutionary potential was

greatly increased through the presence of sensitive CRISPR-KO strains on which the phage could replicate (Supplementary Figure 1) and the lack of detectable coevolution over 30 days when sensitive bacteria were added at each transfer to populations of *P. aeruginosa* and DMS3vir (Chabas *et al.* 2016) suggests that the reduction in mean infectivity that results from the high levels of spacer diversity will prevent local adaptation.

In natural populations, the levels of spacer diversity that can be generated by CRISPR are dependent on at least two factors. First, the rate of spacer acquisition is increased in CRISPR-Cas systems that are 'primed' (Datsenko *et al.* 2012; Fineran & Charpentier 2012; Swarts *et al.* 2012). Priming refers to the situation where a pre-existing spacer in the CRISPR array has partial complementarity with a phage, and this results in the accelerated addition of novel spacers from that phage. Consequently, primed CRISPR-Cas systems can generate high levels of diversity very rapidly, and are better able to deal with phages that are genetically similar to previously-encountered phages compared to completely novel ones. The strain PA14 used in this study is primed against phage DMS3vir, which explains why the WT populations in the coevolution experiment were able to quickly and efficiently generate enough spacer diversity to prevent phages from becoming locally adapted (Figure 2). Similar high levels of spacer diversity have been observed in other natural populations. For example, in studies of *Leptospirillum* biofilms, the trailer end of the CRISPR array was found to be unique at the species level, while the middle of the array was unique at the strain level, and the leader end (where new spacers are incorporated) was found to be unique at the individual level (Tyson & Banfield 2008). Second, CRISPR spacer diversity can also depend on the size of the protospacer-adjacent motif (PAM), which is a conserved sequence of nucleotides flanking each protospacer on the phage genome (Deveau *et al.* 2008). The length of the PAM varies between CRISPR-Cas systems (Mojica *et al.* 2009), and longer PAM sequences reduce the number of putative protospacers that can be sampled from a phage genome, and hence reduce maximum diversity levels.

Previous metagenomic studies have showed that CRISPR spacers tend to match sympatric (but not allopatric) phage genomes (Berg Miller *et al.* 2012; Emerson *et al.* 2013; Held & Whitaker 2009; Kunin *et al.* 2008; Sorokin *et al.* 2010) which

would suggest local adaptation of hosts. However, in such cases it is likely that local adaptation is occurring at an interspecific level (i.e. the phages are of different species), rather than as a direct consequence of coevolution between the same species following different trajectories in different environments. The importance of CRISPR-Cas in such interspecific local adaptation awaits further experimental study.

In summary, it is often assumed that specificity in host-parasite interactions leads to local adaptation due to the selection pressures imposed on parasites to infect the most common host genotypes. In this study we have found that phage are able to locally adapt to their bacterial hosts, but only when the diversity of CRISPR spacers within the host population is low. We therefore propose that even under conditions of high infection specificity, coevolution is unlikely to lead to local adaptation if the host population contains a diverse range of resistance alleles, and this may help to explain why parasite local adaptation is not frequently seen in natural systems (Greischar & Koskella 2007; Kaltz & Shykoff 1998).

Figures

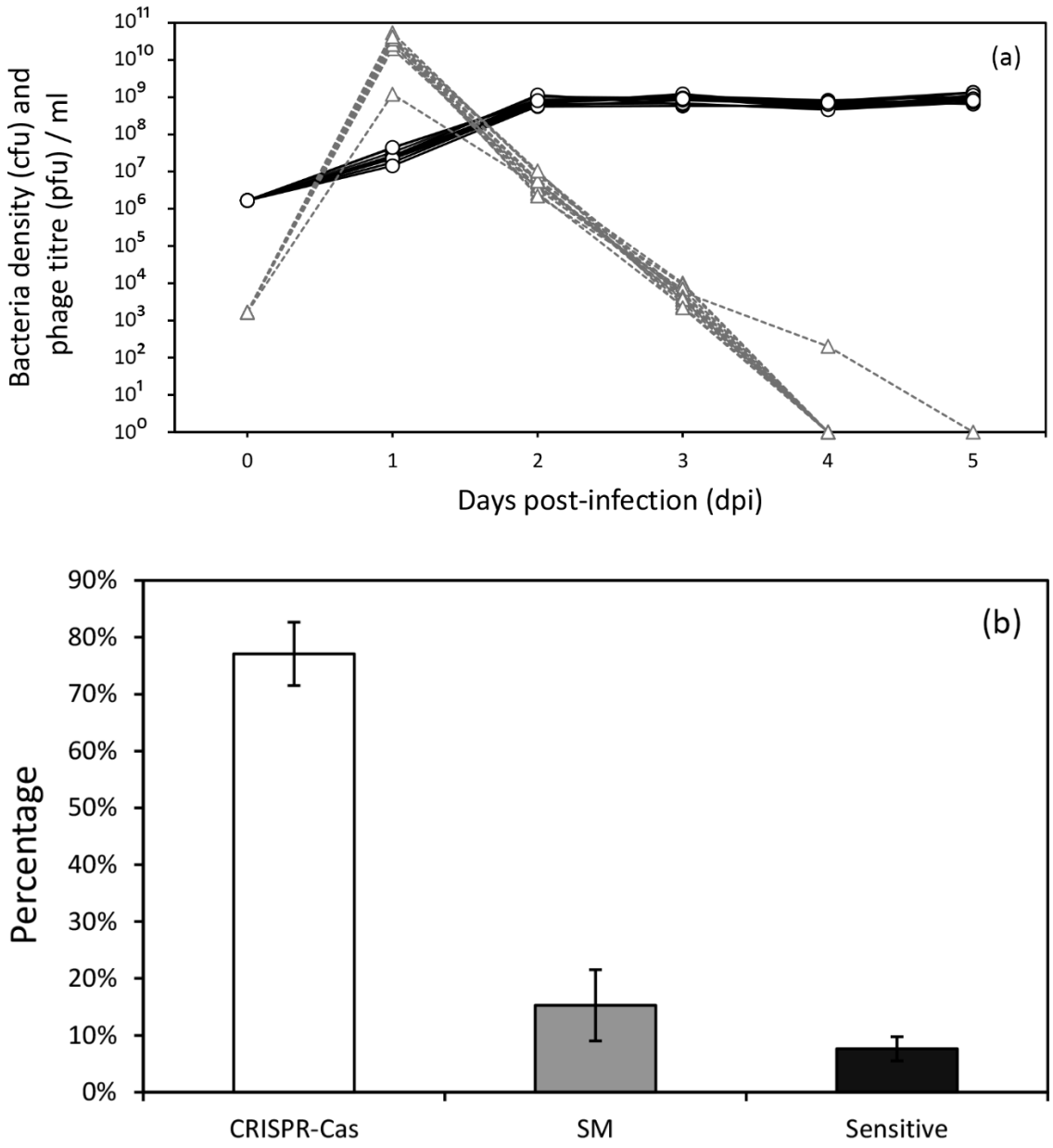


Figure 1. Coevolution of *P. aeruginosa* and DMS3vir over five days post-infection (dpi). **(a)** Bacterial densities (cfu ml^{-1}; solid black lines) and phage titres (pfu ml^{-1}; dashed grey lines). Each line represents an independent replicate (n = 12). The limit of phage detection is 200 pfu ml^{-1}. **(b)** Percentage of bacteria that had evolved CRISPR-Cas (white bar), surface modification (grey bar) or remained sensitive (black bar) to phage at 3 dpi. Values are means of 24 clones from each of the 12 replicates. Error bars indicate 95% confidence intervals (CI).

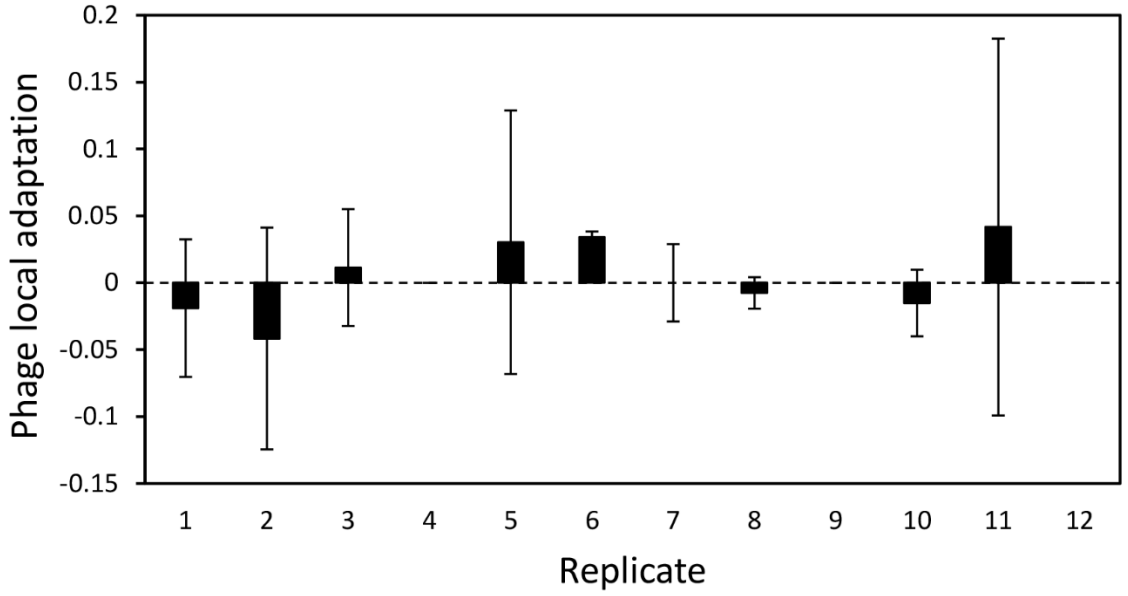


Figure 2. Mean phage local adaptation (foreign vs. local: allopatric phage performance subtracted from sympatric phage performance) for each of the 12 independent replicates at 3 dpi, as measured by cross-streak assays on 24 individual colonies from each replicate. Error bars indicate 95% CI.

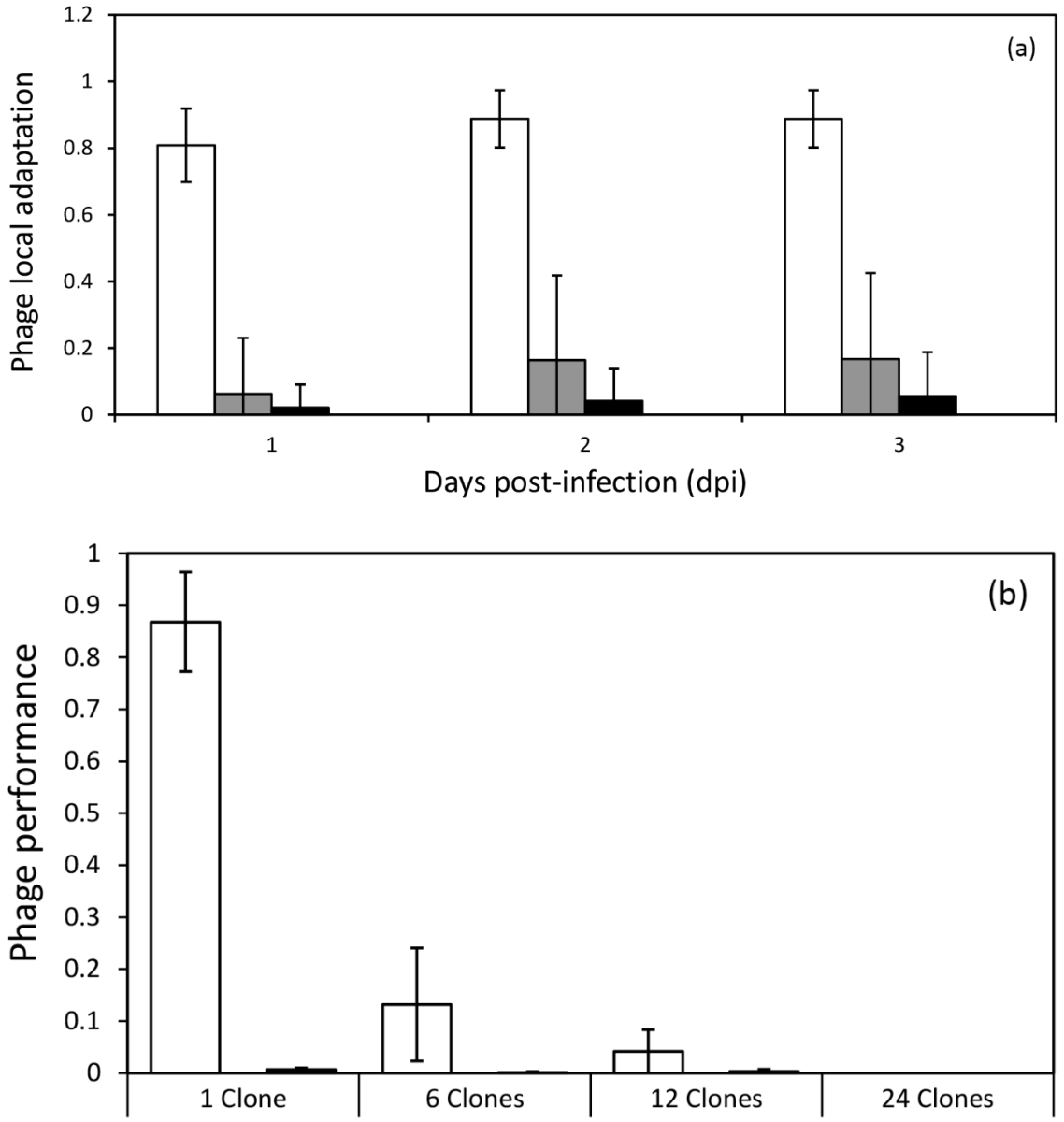


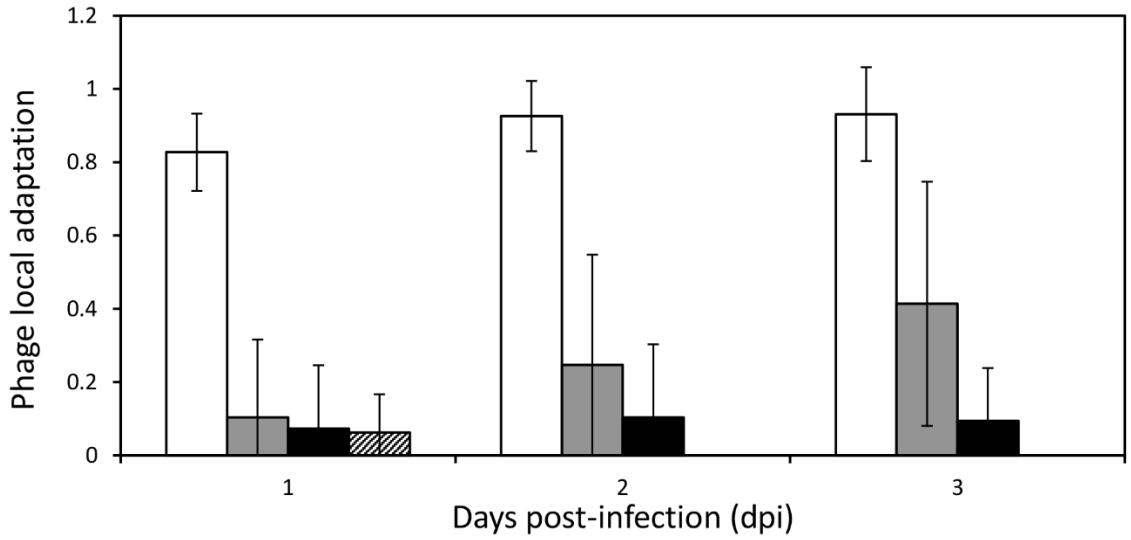
Figure 3. (a) Mean phage local adaptation (foreign vs. local: allopatric phage performance subtracted from sympatric phage performance) on populations with varying CRISPR spacer diversity competed with a surface mutant (*sm*) population at 1, 2 and 3 dpi. White bars: 1 spacer (n = 48), grey bars: 6 spacers (n = 8), black bars: 12 spacers (n = 4), 24 spacers (n = 2; not visible as all values are zero). Error bars indicate 95% CI. **(b)** Mean phage performance (as measured by infectivity) averaged over 1, 2 and 3 dpi on host populations of varying diversity. White bars: sympatric phage, black bars: allopatric phage. Error bars indicate 95% CI.

Tables

	Foreign vs. local				Home vs. away			
Population	LA	CI	T	P	LA	CI	T	p
1-clone CRISPR x *sm*	0.86	0.09	19.1301	<0.0001	0.86	0.09	19.0877	<0.0001
6-clone CRISPR x *sm*	0.13	0.11	2.2851	0.0281	0.13	0.11	2.2049	0.0316
12-clone CRISPR x *sm*	0.04	0.04	n/a	n/a	0.04	0.04	n/a	n/a
24-clone CRISPR x *sm*	0	0	n/a	n/a	0	0	n/a	n/a
1-clone CRISPR x CRISPR-KO	0.89	0.07	24.1288	<0.0001	0.89	0.07	24.3424	<0.0001
6-clone CRISPR x CRISPR-KO	0.25	0.09	5.2190	0.0006	0.25	0.09	5.0784	0.0007
12-clone CRISPR x CRISPR-KO	0.09	0.06	n/a	n/a	0.09	0.06	n/a	n/a
24-clone CRISPR x CRISPR-KO	0.02	0.02	n/a	n/a	0.02	0.02	n/a	n/a

Table 1. Summary of mean phage local adaptation (LA) and 95% confidence intervals (CI) on host populations with varying CRISPR spacer diversity. Foreign vs. local measures the performance of allopatric phages subtracted from the performance of the sympatric phage for each host population. Home vs. away measures the performance of a phage population on sympatric hosts, minus the performance on allopatric hosts.

Supplementary Figures



Supplementary Figure 1. Mean phage local adaptation (foreign vs. local: allopatric phage performance subtracted from sympatric phage performance) on populations with varying CRISPR spacer diversity competed with a functional CRISPR-KO population at 1, 2 and 3 dpi. White bars: 1 spacer (n = 48), grey bars: 6 spacers (n = 8), black bars: 12 spacers (n = 4), dashed bars: 24 spacers (n = 2). Error bars indicate 95% CI.

Chapter 3

CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage

Abstract

CRISPR-Cas is an adaptive prokaryotic immune system that prevents phage infection. By incorporating phage-derived “spacer” sequences into CRISPR loci on the host genome, future infections from the same phage genotype can be recognised and the phage genome cleaved. However, phage can escape CRISPR degradation by mutating the sequence targeted by the spacer, allowing them to re-infect previously CRISPR-immune hosts, and theoretically leading to coevolution. Previous studies have shown that phage can persist over long periods in populations of *Streptococcus thermophilus* that can acquire CRISPR-Cas immunity, but it has remained less clear whether this coexistence was due to coevolution, and if so, what type of coevolutionary dynamics were involved. In this study, we performed highly replicated serial transfer experiments over 30 days with *S. thermophilus* and a lytic phage. Using a combination of phenotypic and genotypic data, we show that CRISPR-mediated resistance and phage infectivity coevolved over time following an arms race dynamic, and that asymmetry between phage infectivity and host resistance within this system eventually causes phage extinction. This work provides further insight in the way CRISPR-Cas systems shape the population and coevolutionary dynamics of bacteria-phage interactions.

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats and their associated cas genes (CRISPR-Cas) form an adaptive immune system that is found in approximately 50% of all bacteria and 90% of archaea (Grissa *et al.* 2007). CRISPR-Cas confers immunity to phage infection by incorporating phage-derived sequences into CRISPR loci on the host genome. These loci consist of repeating sequences ("repeats") that are interspaced by sequences ("spacers") derived from phage and other mobile genetic elements of typically around 30 nt in length. RNA transcripts of CRISPR loci are processed and form a ribonucleoprotein complex with Cas proteins that can recognise and cleave complementary nucleic acid sequences, preventing future infections by the same phage genotype. CRISPR-Cas systems are highly diverse, and are currently ordered into two classes, six types and 33 subtypes based on their *cas* gene composition, gene synteny and CRISPR repeat sequences, with clear differences in the molecular mechanisms of different variants (Koonin *et al.* 2017).

In some natural environments, bacteria with CRISPR-Cas systems appear to coevolve with phage over long time periods (Andersson & Banfield 2008; Laanto *et al.* 2017). However, studying the dynamics of these coevolutionary interactions under controlled laboratory conditions has been limited by the availability of adequate model systems. Specifically, while many bacteria encode CRISPR-Cas immune systems, under laboratory conditions the vast majority do not evolve CRISPR-based immunity upon phage or plasmid infection, or do so at such low frequencies that they are detectable only with deep-sequencing approaches. Such low-frequency CRISPR evolution is unlikely to significantly contribute to the reciprocal selection between the bacteria and the phage. Currently, only two bacterial species have been found to naturally evolve (almost) exclusively CRISPR-based immunity under laboratory conditions: *Streptococcus thermophilus* strains DGCC7710 and LMD-9 (Deveau *et al.* 2008; Horvath *et al.* 2009; Horvath *et al.* 2008), and *Pseudomonas aeruginosa* strain UCBPP-PA14 (Westra *et al.* 2015).

Early studies with *S. thermophilus* demonstrated that phage can overcome CRISPR immunity by evolving point mutations in the sequence targeted by the

spacer (the “protospacer”), or in the protospacer-adjacent motif (PAM) (Deveau *et al.* 2008), a conserved sequence immediately adjacent to the protospacer that is used by the bacteria to discriminate between self (i.e. CRISPR loci) and non-self (i.e. phage) DNA (Mojica *et al.* 2009; Semenova *et al.* 2011; Westra *et al.* 2013). This observation suggested a possible scenario for coevolution in free-running systems, where bacteria acquire spacers over time and phage escape via point mutations in the corresponding protospacers or PAMs (Childs *et al.* 2012; Iranzo *et al.* 2013; Weinberger *et al.* 2012). Consistent with this idea, it was reported that *S. thermophilus* can coexist with phage over many generations, and that for each treatment the single experimental population displayed large fluctuations in its spacer repertoire and an increase in the frequency of point mutations in phage genomes over time (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015; Sun *et al.* 2016). However, a more recent study suggested that coevolution is unable to explain long-term coexistence of *S. thermophilus* and its phage, suggesting instead that this may be driven by back mutation of bacteria with CRISPR immunity to sensitive phenotypes (Weissman *et al.* 2018), which would provide a continuous supply of sensitive hosts for phage to amplify in. Such loss of CRISPR immunity due to mutation has also been observed at high frequencies in *Staphylococcus epidermidis* (Jiang *et al.* 2013), and reversion to sensitive phenotypes more generally may be an important mechanism for bacteria-phage coexistence (Chaudhry *et al.* 2018).

Given that phage and bacteria with CRISPR-Cas have been found to coexist (Paez-Espino 2013, 2015) and coevolve (Andersson & Banfield 2008; Laanto 2017), we set out to discover the specific role, if any, that CRISPR plays in this coevolution. We performed highly replicated, long-term (30-day) serial transfer experiments with *S. thermophilus* and its lytic phage 2972. Our phenotypic assays demonstrate that bacteria and phage coevolved in these experiments during at least the first 9 days (approximately 70 generations). We next examined the type of coevolutionary dynamics during this period, with a clear distinction between fluctuating selection dynamics (FSD), where rare host and pathogen genotypes are favoured through negative frequency-dependent selection, and arms race dynamics (ARD), where host resistance and phage infectivity increase over time (Vale & Little 2010). We found that CRISPR-mediated immunity and phage infectivity increase over time. Further, our genotypic data show that

patterns of resistance and infectivity were explained by bacteria acquiring novel spacers against the phage, and the phage evolving mutations in the regions targeted by the spacers. Collectively, our results provide the first experimental evidence that CRISPR-phage coexistence is caused by arms-race coevolution in this model system.

Methods

Strains used in the study

We used the lactic acid bacterium *Streptococcus thermophilus* DGCC7710 wild-type (WT) and its lytic phage 2972 as a model system. DGCC7710 has four CRISPR-Cas systems, two of which (CRISPR1 and CRISPR3) are active during infection with phage 2972 and both are classified as Type II-A (Carte *et al.* 2014; Horvath *et al.* 2008).

Phage 2972 amplification

An overnight culture of *S. thermophilus* was transferred 1:10 into fresh LM17 medium (M17 broth supplemented with 0.5% α-Lactose) containing 10 mM $CaCl_2$ and incubated shaking at 180 revolutions per minute (rpm) at 42˚C. When the culture reached log phase (OD_{600} ~ 0.25) approximately 10^6 plaque forming units (pfu) of phage 2972 were added and the culture was incubated under the same conditions for two hours, at which point cells had fully lysed. Lysates were centrifuged and filtered through a 0.22 µm filter, and the resulting phage stocks were stored at 4˚C.

Long-term co-culture experiment

Prior to commencing the experiment, *S. thermophilus* was acclimatised in LM17 medium at 42˚C and 180 rpm for two days, with a 1:100 transfer into fresh LM17 after 24 hours. To start the co-culture experiment, bacteria were transferred 1:100 into LM17 media supplemented with 10 mM $CaCl_2$ and infected with either 10^9, 10^8, 10^7 or 10^6 pfu of phage 2972, with 12 independent replicate experiments per treatment, followed by incubation at 42˚C while shaking at 180 rpm. Replicates were transferred 1:100 into fresh LM17 + 10 mM $CaCl_2$ every 24 hours and phage titres and bacterial densities were measured every 24 hours for a period of 30 days, or until no phage was detected for four consecutive days. Bacterial densities were determined through plating and colony counts, while phage densities were measured by plaque assays. These were performed by mixing phage dilutions with WT bacteria in soft agar overlays (LM17 + 10 mM $CaCl_2$ and 0.5% agar), poured onto hard agar (LM17 + 10 mM $CaCl_2$ and 1.5% agar).

Phage survival

Phage survival and mean time to extinction over the course of the experiment were analysed using a Cox proportional hazards model from the survival package (Therneau & Lumley 2015).

Measuring the evolution of infectivity and resistance

To measure whether host resistance and phage infectivity evolved during their co-culture, we isolated phage clones and bacterial clones from the treatment where bacteria were infected with 10^8 pfu phage. For this analysis we used eight out of 12 replicate experiments from this treatment where phage persisted for at least nine days. We chose nine days as the maximum time point for analysis because after this point there would not have been enough surviving replicates for meaningful statistical power. Day four was also chosen for analysis as this was the midway point. Phage extracted from 1, 4 and 9 days post-infection (dpi) were subjected to plaque assays as described above. For each replicate and time point, twelve plaques were randomly picked and amplified in 96 well plates containing LM17 + 10 mM $CaCl_2$ in which WT bacteria were inoculated 1:100 from a fresh overnight culture. Bacteria extracted from the same time points were diluted and plated overnight, and twelve colonies from each replicate were picked at random and used to make soft agar overlays on LM17 agar lawns. To examine the evolution of phage infectivity for each of the eight replicates, the 36 phage clones that were isolated (12 phage clones x 3 time points) were spotted onto 36 bacterial lawns corresponding to the bacterial clones isolated from the same replicate (i.e. 12 bacterial clones x 3 time points). Phage were classified as being infective against a particular bacterial clone if a plaque was visible on the lawn after incubation at 42˚C for 24 hours. If no plaque was visible, the host was classified as resistant.

Using the data from the experiments described above, we measured the evolution of phage infectivity as the proportion of bacterial clones that phage from each time point from the same replicate experiment could infect (i.e. how phage infectivity range changed over time). In a similar way, we measured the evolution of host resistance as the proportion of all phage genotypes from the same replicate experiment that could be resisted by bacteria from each time point (i.e. how host resistance range changed over time). Infectivity or resistance was analysed in a Generalised Linear Model (GLM) with genotype as a fixed effect

and a binomial family with a logit link function. Mean infectivity or resistance was then analysed for each time point in a Generalised Linear Mixed Model (GLMM) using the lme4 package (Bates *et al.* 2014), with time point as a fixed effect and replicate as a random effect. Model coefficients and confidence intervals were transformed from logits to probabilities prior to presentation.

Time-shift experiment

Because the susceptibility and resistance of bacterial clones to phage from past, present or future time points was determined (Table 1), our phenotypic assay also served as a time-shift experiment (Gaba & Ebert 2009). Time-shift experiments involve challenging samples of host or pathogen populations from a particular time point against samples of pathogen or host populations from contemporary, past and future host or pathogen populations. Time-shift experiments are a powerful tool to characterise underlying coevolutionary processes, and have been used in several host-pathogen systems (Gandon *et al.* 2008; Koskella & Lively 2007), including bacteria-phage (Buckling & Rainey 2002; Hall *et al.* 2011; Koskella 2014). Our phage infectivity and host resistance data was analysed as a time-shift experiment by first scoring each pairwise challenge as 'past', 'present', or 'future', with reference to the phage's background compared to the host. Infectivity was then analysed in a GLMM with phage background as a fixed effect and phage genotype as a random effect. Models had a binomial family with a logit link function.

To test for the relative importance of arms race dynamics (ARD) versus fluctuating selection dynamics (FSD), we estimated the strength of the Genotype x Environment (G×E) effect on infectivity and resistance following Hall *et al.* (2011). Under a simple arms race, all hosts should be more susceptible to phage from their future compared to their past or present, independent of genotype. Environment (E) therefore refers to the time point from which phage originate in pairwise challenges. By contrast, under fluctuating selection different host genotypes will vary in their susceptibility to hosts from their past, present or future. Measuring which proportion of the variation in susceptibility across phage environments can be explained by the interaction between the environment and host genotype (G) can therefore be used to measure the relative contribution of FSD. Increasing values of this proportion (G×E / E) relate to increasing

differences in susceptibility among host genotypes. We estimated this by calculating the ratio of the mean square (MS) of an Environment-only model to the MS of a G×E model for each replicate at each time point. These ratios were then analysed in a GLMM with time point as a fixed effect and replicate as a random effect, with a normal family and square root link function.

Statistical analyses

For all experiments, statistical analyses were carried out in R v3.5.0 (R Core Team 2018), and graphics were generated using r-base and the ggplot2 package (Wickham 2009). Model selection followed a nested design, and the final models in all analyses were selected based on the reduction of heteroskedasticity, χ^2 tests, and AIC comparisons (Akaike 1973; Burnham & Anderson 2003, 2004). Where appropriate, tests were Bonferroni adjusted using the multicomp package (Hothorn *et al.* 2008).

Spacer sequence analysis

For all bacterial clones that were isolated from the eight replicate experiments where bacteria had been infected with 10^8 pfu of phage 2972, expansion of the CRISPR1 and CRISPR3 arrays was analysed using PCR to determine whether spacer acquisition had taken place (12 clones x 3 time points x 8 replicates = 288 clones in total). Clones that had acquired new spacers were further analysed by Sanger sequencing of the amplicon (Source Bioscience, UK), followed by mapping of the spacers against the phage 2972 genome (Accession: NC 007019.1) using BLAST followed by manual verification with Geneious v9.1.8 (Kearse *et al.* 2012). Spacer diversity was calculated as the pairwise difference (PWD) among nucleotides between spacer sequences. The effect of spacer number and diversity on infectivity was analysed in a GLMM, with either number or diversity as fixed effects and replicate as a random effect. Models had a binomial family and logit link function.

Phage sequence analysis

To understand whether phages could escape CRISPR immunity through target site (protospacer) mutation, we selected phage clones and sequenced the protospacer(s) that would match the spacer(s) present among the 12 isolated bacterial clones in a replicate and analysed the protospacers and their associated

protospacer-adjacent motifs (PAMs) for SNPs that could explain the ability of the phage to overcome CRISPR immunity. For phage clone selection we used the individual infectivity matrices (Supplementary Table 1), and we only sequenced phage clones from infection matrices that showed phage infectivity on hosts that had acquired spacers i.e. excluded from analysis were phages from those matrices where none of the 12 host clones of a replicate had acquired spacers or where none of the 12 phage clones of a replicate showed infectivity. Sequenced phage clones were taken from 1, 4 and 9 dpi. At least two individual phage clones were selected from each matrix that was analysed, based on their ability to infect CRISPR-immune hosts. When an infection matrix showed a high degree of variation between phage clones, more than two phage clones were analysed so that most of the variation in infectivity would be covered (e.g. in the matrix of replicate 7, T9 phage:T9 host, phage clones 1, 2, 5, 6 , 8, and 9 were selected [Supplementary Table 1]). Where possible, one phage clone that did not show infectivity to any of the 12 bacterial clones from each matrix was taken along to serve as a control for protospacer sequencing alongside an ancestral phage. Primers for protospacer sequencing were designed in Geneious using the available spacer information (Supplementary Table 2), and a total of 51 phage clones were sequenced using direct PCR on the chloroformed phage stocks.

To identify point mutations, sequences were first mapped to the 2972 genome using Geneious v9.1.8. Single nucleotide polymorphisms (SNPs) in either the seed sequence or PAM were identified (Deveau *et al.* 2008; Horvath *et al.* 2008), and SNP locations were then compared against the protospacer sequence targeted by the CRISPR array of each clone that phage had been challenged against (Supplementary Table 3). Phage with SNP(s) in the seed sequence or the PAM of the targeted protospacers were scored as ‘predicted infective’. We found that approximately 70% (241 / 348) of predicted infectious phage were measured as successfully infecting a host (Table 3). The remaining predicted infections that were not measured as successful may be attributable to partial CRISPR resistance, or other resistance mechanisms such as surface modification (*sm*), although *sm* has not been observed in *S. thermophilus* (Sylvain Moineau, personal correspondence). There is also likely to be some degree of experimental error in our assay. Using data from the phenotypic assay, we then analysed the effect of mutation on infectivity. The effect of escape by point

mutation on infectivity was modelled in a GLMM as the proportion of infections associated with phage that had a SNP in the protospacer seed sequence or PAM. We analysed the effect of the evolution of the number of SNPs in all targeted sequences matching the host's CRISPR array by first subtracting the number of targeted protospacers that had evolved from the total number of spacers in each host. This gives the number of targeted sequences that had not evolved. The proportion of infections was then modelled against these values. All models included replicate as a random effect and used a binomial family with a logit link function.

Results

We set out to first examine the generality of the previously reported population dynamics following infection of *S. thermophilus* DGCC7710 with a single phage 2972. We therefore infected 12 replicate experimental populations of *S. thermophilus* DGCC7710 with either 10^6, 10^7, 10^8 or 10^9 plaque forming units (pfu) of phage 2972 (12 independent replicates per treatment; 48 populations in total), and monitored the bacterial and phage population densities on a daily basis for 30 days. For the first three days following infection, phage titres remained fairly constant in most replicates between 10^6-10^8 pfu ml^{-1}, with the exception of the highest phage treatment (10^9 pfu) where phage and bacteria went extinct in 11 out of 12 replicates (Figure 1). Lower phage titres were correlated with higher host densities (z = -0.31, 95% CI = -0.42, -0.19, $p < 0.0001$). With the exception of the 10^9 treatment, this relationship between phage and host titres was the same among treatments ($F_{2,551} = 2.24$, $p = 0.11$). At 16 days post-infection (dpi), the phage had gone extinct in 44 / 48 replicates, and phage persisted for the entire 30-day duration of the experiment in two replicates, one each in the 10^7 and 10^8 pfu treatments. For the treatments where bacteria survived, the mean time until phage extinction in days was as follows: for the 10^9 pfu treatment: 2 ± 0.54 days; 10^8 treatment: 11.50 ± 1.77 days; 10^7 treatment: 11.50 ± 2.12 days; and 10^6 treatment: 7.67 ± 1.67 days (mean ± standard error).

Using these experimental lines, we first determined whether the coexisting bacteria and phage had evolved during their co-culture. Since the population dynamics associated with the 10^6-10^8 pfu infection regimes was virtually identical, we decided to limit our downstream analyses to the 10^8 treatment only. Further, to achieve sufficient power in our analyses, we selected replicates where bacteria and phage coexisted for at least nine days, resulting in a total of eight replicate populations that were examined in detail (Figure 1). We then isolated 12 bacterial clones and 12 phage clones from each replicate at 1, 4 and 9 days post-infection (dpi). Using the 288 phage and 288 bacterial isolates, we first examined whether the phage and bacteria had evolved increased infectivity and resistance over time. This was done by measuring the resistance of each individual bacterial clone against all phage clones derived from the same replicate, and measuring the infectivity of each individual phage clone against all bacterial clones from the same replicate. This analysis revealed that mean phage infectivity (the proportion

of all host genotypes that can be infected by a given phage genotype) increased significantly from 0.29 (CI = 0.08, 0.48) at 1 dpi to 0.57 (CI = 0.37, 0.74) at 4 dpi, but remained stable at 0.53 (CI = 0.33, 0.74) from 4 to 9 dpi. Mean host resistance (the proportion of all phage genotypes resisted by a given host genotype) increased significantly each time point, from 0.01 (CI = 0.00, 0.05) at 1 dpi to 0.67 (CI = 0.18, 0.96) at 4 dpi, and to 0.99 (CI = 0.96, 0.99) at 9 dpi (Figure 2). Collectively, these data show that bacteria evolved to resist essentially all phage genotypes by 9 dpi, but phage did not evolve high levels of infectivity to match.

Having established that bacteria evolved increasing resistance and that phage evolved increasing infectivity over time, we next examined whether both species coevolved; and if so, what type of coevolutionary dynamics were associated with this system. To answer this question we performed a phenotypic time-shift experiment whereby bacteria were exposed to phage from their past, present and future (Gaba & Ebert 2009; Koskella 2014) which enabled us to measure infectivity and resistance patterns over time. Because individual genotypes may differ in their response to time-shift challenges, generalised linear mixed models (GLMMs) with phage genotype as a random effect and phage background (Table 1) as a fixed effect were used to analyse time-shift data. This analysis showed that the original time-point of the phage with respect to the host had a significant effect on infectivity ($\chi^2_{(4,10044)} = 5.35$, $p < 0.0001$, $R^2 = 0.39$). Hosts were least susceptible to infection from past phage, more susceptible to contemporaneous phage, and most susceptible to phage from their future (Figure 3a). This pattern of increasing susceptibility from past to future phage generally held true when each pairwise combination of host and phage time point was considered (Table 2 & Supplementary Figure 1). Finally, host susceptibility to phage from the same time point declined consistently from 1 to 9 dpi (Figure 3b). These data are consistent with an arms race dynamic (ARD) where hosts and pathogens escalate resistance or infectivity over time, but one in which host resistance eventually outpaces pathogen infectivity.

We formally tested for the relative importance of arms race versus fluctuating selection in our experiment by estimating the strength of the genotype x environment (G×E) effect on infectivity and resistance (Hall *et al.* 2011). Stronger G×E effects are consistent with stronger fluctuating selection (see *Methods*). This

analysis showed that variation among genotypes was weak, consistent with a limited G×E effect (Supplementary Figures 2a & 2b). The strength of the G×E effect did not change significantly with respect to time point for either phage infectivity ($\chi^2_{(2,24)} = 1.93$, $p = 0.38$, $R^2 = 0.13$) or host resistance ($\chi^2_{(2,24)} = 1.46$, $p = 0.48$, $R^2 = 0.11$). Collectively, these data demonstrate that *S. thermophilus* DGCC7710 and phage 2972 coevolved under these experimental conditions, and that the dynamics of their coevolution predominantly follows an arms race.

Based on previous studies (Deveau *et al.* 2008; Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015), we predicted that this ARD was driven by reciprocal adaptation of the hosts' CRISPR array and the phage protospacers it targets. To test this, we first performed PCR analysis on the CRISPR1 and CRISPR3 loci of each bacterial clone to verify that the mechanistic basis of resistance was in fact due to the acquisition of novel CRISPR spacers. This revealed that the mean number of spacers per clone increased over time ($\chi^2_{(6,1140)} = 32.9$, $p < 0.0001$) (Figure 4a), and that all clones had acquired at least one spacer by 9 dpi (M = 0.55, CI = 0.45, 0.65) (Figure 4b). Host resistance increased significantly with the number of acquired spacers (C = 1.91, z = 17.22, $p > 0.0001$, $R^2 = 0.66$) (Supplementary Figure 3a). These data demonstrate that all clones that had acquired resistance had also acquired at least one novel spacer in either CRISPR1 or CRISPR3, suggesting that resistance is CRISPR-mediated. Further, Sanger sequencing of all CRISPR amplicons confirmed that all spacers that had been acquired indeed targeted the phage 2972.

Using these sequencing data we determined the level of spacer diversity that naturally evolved within each replicate, since this is an important determinant of CRISPR-phage coevolution (Childs *et al.* 2014; van Houte *et al.* 2016b). Consistent with deep sequencing analyses of previous co-culture experiments (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015), our data showed that spacer diversity, measured as the pairwise difference (PWD, with 0 indicating all spacers were shared among clones, and 1 indicating all spacers were unique) among spacer sequences, was generally low (grand mean = 0.25). Despite this, there was clear qualitative variation in spacer diversity between different replicates (Figure 4c). Mean CRISPR genotype richness – the number of different CRISPR alleles we detected - was also low, but increasing, across the sampled time points

(1 dpi = 1, 4 dpi = 1.5, 9 dpi = 2.25). The diversity patterns become especially apparent when the spacers are mapped against the phage genome (Figure 5) which shows that the spacer composition between time points can change completely, suggestive of selective sweeps of the population. Consistent with previous theory and data (Childs *et al.* 2014; van Houte *et al.* 2016b) we found that host resistance increased with sequence diversity, in terms of PWD ($C = 5.26$, $z = 0.27$, $p < 0.0001$, $R^2 = 0.44$) (Supplementary Figure 3b).

We next tested the hypothesis that the coevolutionary arms race we observed in these experiments was caused by reciprocal adaptation of the phage through the acquisition of point mutations in the target sequences. Such mutations have been observed in a previous co-culture experiment (Paez-Espino *et al.* 2015), and provide a known mechanism for phage to overcome CRISPR resistance (Deveau *et al.* 2008). To examine whether phage infectivity could be explained by the acquisition of point mutations, we first selected from the phenotypic assays a representative number of 56 different phage clones with different infectivity patterns (i.e. covering both infective and non-infective phenotypes). We then PCR amplified their protospacer sequences based on the CRISPR spacer sequence data, followed by Sanger sequencing of the amplicons. This showed that 38 out of 51 selected phage clones had acquired at least one single nucleotide polymorphism (SNP) in the protospacer sequence or PAM (Figure 6a); the majority (33) of which were protospacer mutations (Figure 6b). Crucially, analysis of the infectivity patterns of sequenced phage showed that the mean number of infected hosts was significantly higher when phage had a SNP in the protospacer sequence or PAM compared to phage with no detectable mutations ($\chi^2_{(1,696)} = 32.22$, $p < 0.0001$, $R^2 = 0.31$) (Figure 6c). Finally, phage that had evolved SNPs in all sequences that were targeted by the host's CRISPR array (0 targeted sequences) had significantly higher infectivity compared to phage that carried one or more unmutated target sequences ($\chi^2_{(1,696)} = 59.29$ $p < 0.0001$, $R^2 = 0.74$) (Figure 6d). These data demonstrate that the acquisition of point mutations in the protospacer sequence in response to evolution of CRISPR immunity is the primary mechanism of phage reciprocal adaptation, driving the increase in phage infectivity during the observed arms race dynamic.

Discussion

S. thermophilus DGCC7710 readily evolves CRISPR-based resistance in response to phage 2972 through spacer acquisition in two active CRISPR loci (CRISPR1 and CRISPR3). In return, phage can escape CRISPR immunity by evolving point mutations in the protospacer sequences targeted by CRISPR. This mechanism of host resistance and pathogen infectivity suggests a possible scenario for coevolution, where bacteria acquire spacers over time and phage accumulate escape mutations.

Consistent with earlier work on *S. thermophilus* and phage 2972, we found that phage can coexist with bacteria over many generations despite the presence of CRISPR-based host immunity. Our phenotypic data using bacteria and phages isolated at 1, 4 and 9 days post infection (dpi) demonstrated an arms-race dynamic characterised by increasing host resistance and phage infectivity over time, and hosts being more resistant to phage from the past compared to the present and future. The predominant underlying mechanism of coexistence during this period appears to be reciprocal adaptation of the hosts' CRISPR array and the corresponding phage protospacer sequences. Analysis of hosts' CRISPR arrays show that they readily acquire phage-derived spacers, that hosts acquire more spacers over time, and that host resistance is strongly associated with both spacer acquisition and spacer number. These results, when combined with phenotypic data, showed that phages that had 'escaped' CRISPR were on average more infective, and phages that had evolved SNPs in *all* target sequences matching the host's CRISPR array were most infective compared to those with an incomplete match.

It is notable that while hosts evolve resistance against essentially all phage, this is not matched by similarly broad phage infectivity range; phage at the last time point (9 dpi) could infect just over half of all hosts. The infectivity of contemporary phage also declines with time, suggesting that the evolution of host resistance 'outpaces' that of phage infectivity. This asymmetry between host resistance and phage infectivity is consistent with the idea that bacterial hosts are often 'ahead' in coevolutionary arms races (Buckling & Brockhurst 2012). While asymmetrical arms races in other studied bacteria-phage systems are generally driven by a binary shift to a phage-resistant surface mutant (Lenski & Levin 1985; Westra *et*

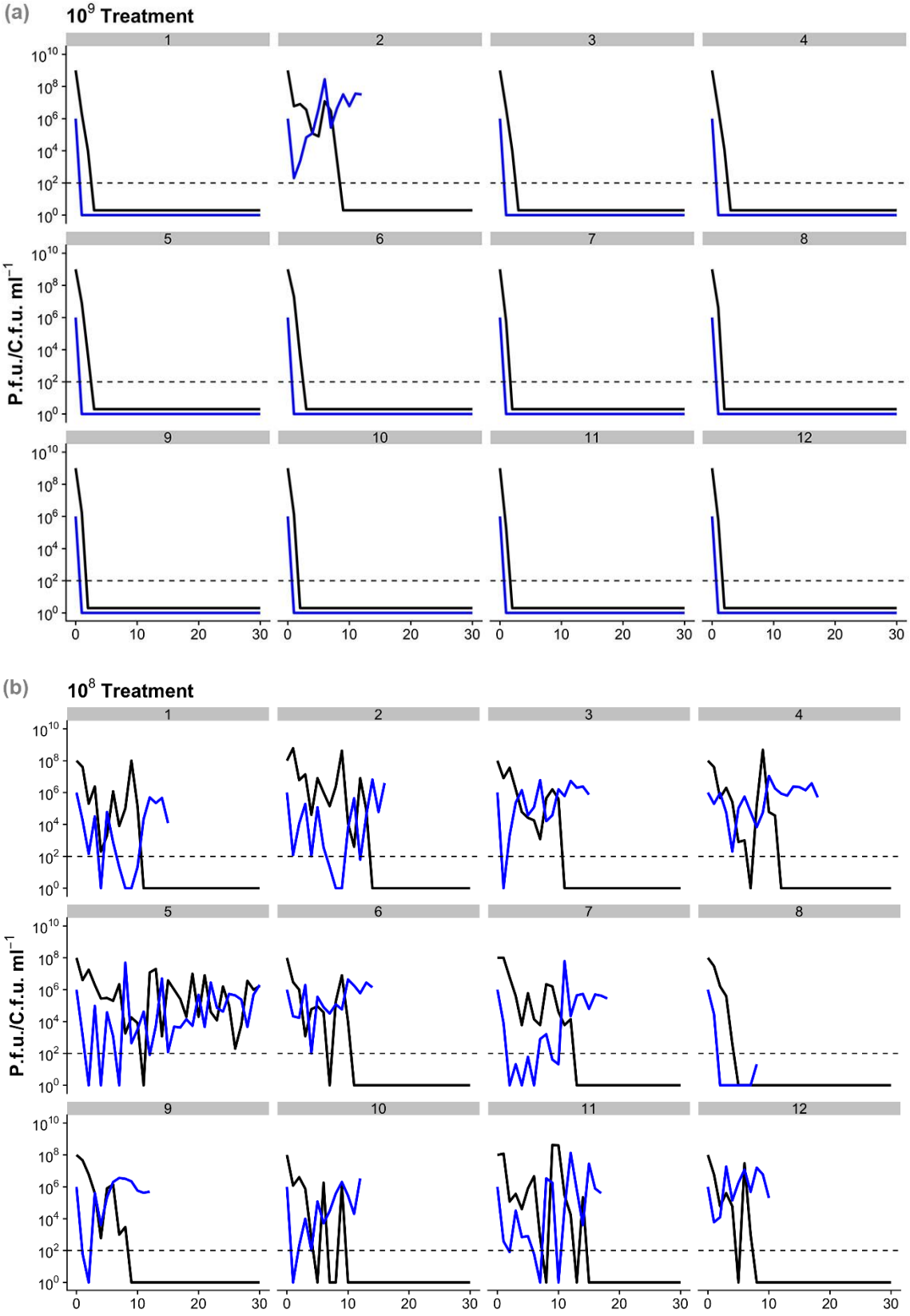
al. 2015), CRISPR-phage interactions suggest an alternative. Hosts can acquire multiple novel spacers with only a marginal cost (Vale *et al.* 2015), but phage mutation is limited by mutation supply (Lenski & Levin 1985; Levin *et al.* 2013). In addition, full phage infectivity requires mutations in all the protospacers targeted by the host CRISPR array, which becomes increasingly difficult when individual hosts and populations acquire a greater number and diversity of spacer sequences over time (Levin *et al.* 2013; van Houte *et al.* 2016b). It is likely that this asymmetry leads to the repeatable phage extinctions we observed.

Interestingly, this and previous studies occasionally found quasi-stable long-term coexistence of bacteria and phage (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015; Sun *et al.* 2016; Weissman *et al.* 2018). Previous work suggests that this may be driven by back-mutation of resistant hosts towards sensitivity (Weissman *et al.* 2018). In the case of *P. aeruginosa* (where phage are unable to coevolve with the host due to the high levels of spacer diversity that naturally evolve) a continuous supply of sensitive hosts can allow for bacteria-phage coexistence (Chabas *et al.* 2016; van Houte *et al.* 2016b; Westra *et al.* 2017).

Our data clearly shows that host genotype diversity richness (i.e. the number of hosts with different CRISPR arrays) increases over time. This data, together with the rapid and repeatable phage extinction after day 9 in our experiment, indirectly supports the idea that *S. thermophilus* hosts receive a synergistic benefit from population-level CRISPR diversity in the context of phage infection. Previous studies of *P. aeruginosa* PA14 and *S. thermophilus* have shown that population-level CRISPR spacer diversity can limit phage persistence (van Houte *et al.* 2016b). Host diversity is known to be a key determinant of pathogen spread and the transmission of disease, as increasing host genetic diversity decreases the chances of a parasite successfully infecting a host, or an infected host passing on an infection to a genetically-similar neighbour (Keesing *et al.* 2010; Lively 2010; King & Lively 2012; Ashby & King 2015). In our experiments, although phage infectivity was still possible even with a partial match to the host's CRISPR array, infectivity was reduced to the extent that it may be sufficient to cause the rapid phage extinctions we observed.

In at least some natural environments, bacteria that evolve CRISPR resistance and the phage they target can coexist (Andersson & Banfield 2008). This may be due to CRISPR-phage coevolution, as recently observed for a fish pathogen and its phage (Laanto *et al.* 2017), but long-term coexistence may also be explained by various other ecological and evolutionary factors that are absent from our simple laboratory environments. For example, previous experiments suggest that longer periods of bacteria-phage coexistence are reached when experimental treatments contained multiple different phages (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015). Further, phage under these conditions were found to escape not only by mutation, but also by recombination (Paez-Espino *et al.* 2015). This is consistent with observations from other natural environments where phage recombinants were correlated with CRISPR activity (Andersson & Banfield 2008). These examples highlight how biotic and abiotic complexities may be key in shaping the ecological and evolutionary dynamics of host-pathogen interactions, which we are only starting to understand in the context of CRISPR-phage interactions.

Figures


(a)
10^9 Treatment
(b)
10^8 Treatment
P.f.u./C.f.u. ml^-1
1
2
3
4
5
6
7
8
9
10
11
12
0
10
20
30

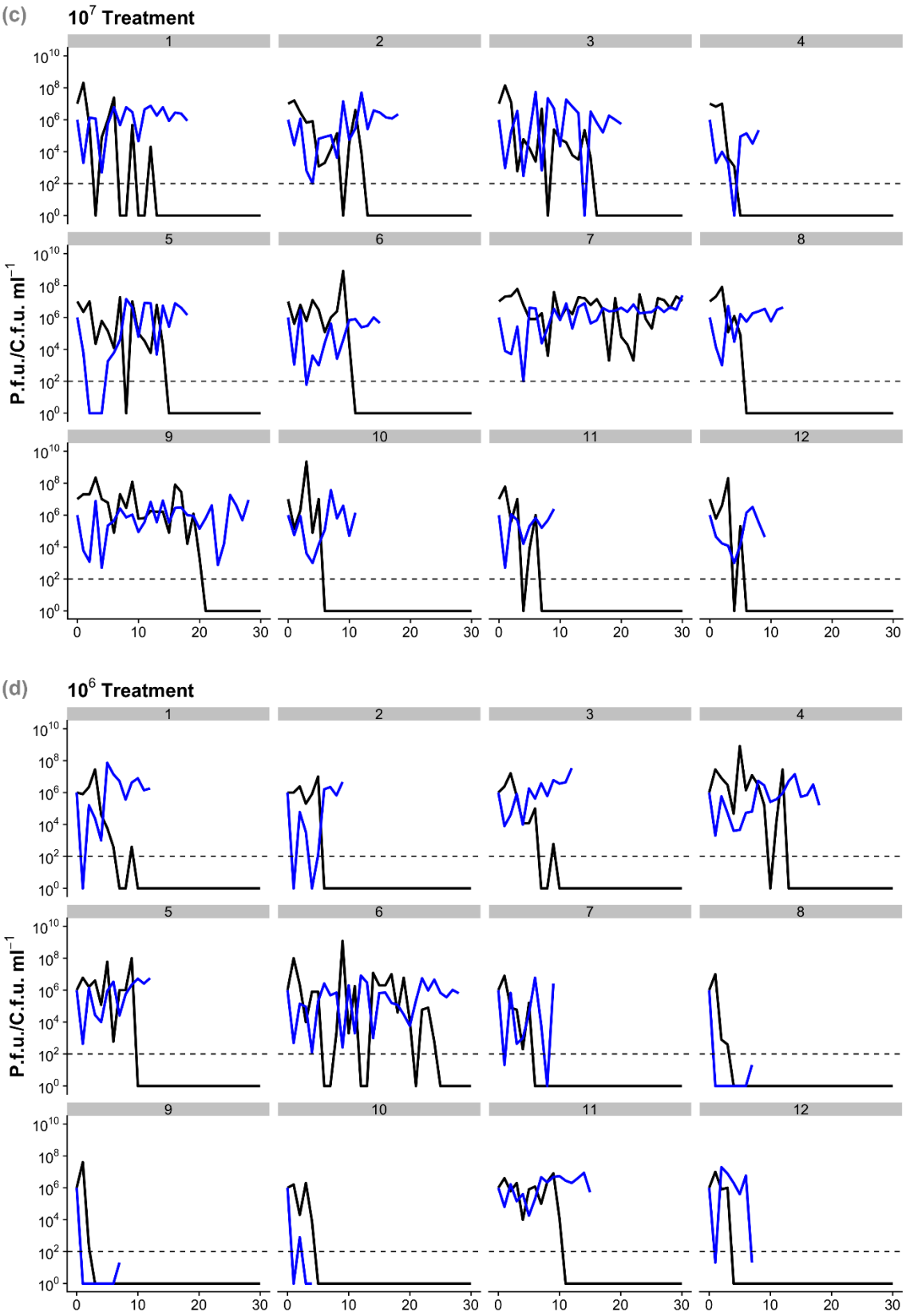


Figure 1. Phage and host population dynamics over time in each replicate**. (a-d)** 10^9-10^6 pfu phage treatments, respectively, with replicate identity indicated above each sub-panel. Phage titres (plaque-forming units; pfu ml^{-1}) are shown in black and host densities (colony-forming units; cfu ml^{-1}) are shown in blue. The level of detection is 200 pfu ml^{-1} (dashed line).

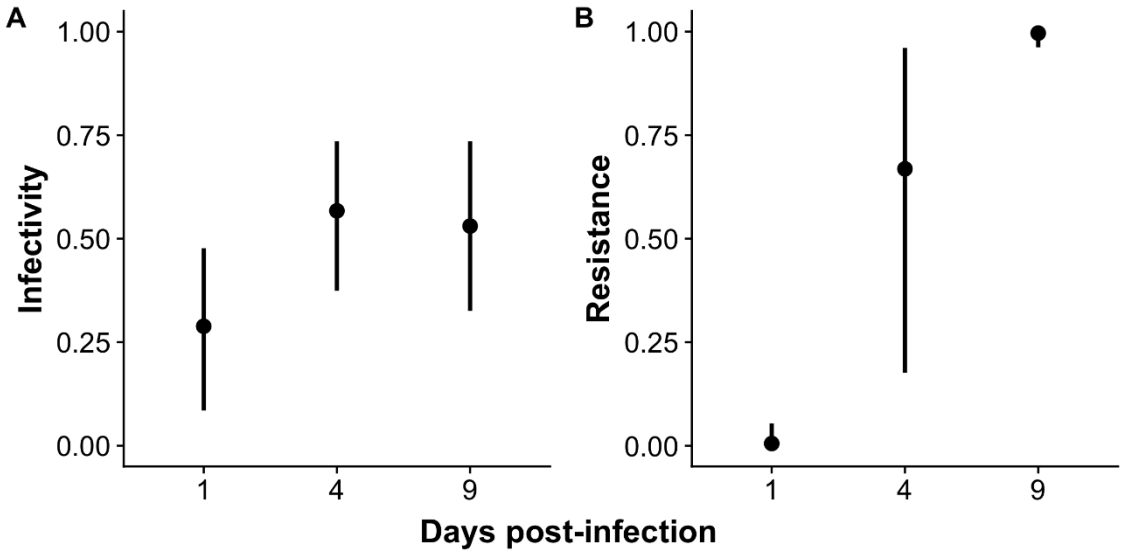


Figure 2. Evolution of infectivity and resistance over time. **(a)** Phage infectivity over time, represented as the proportion of host genotypes from all time points that were infected by a given phage genotype in each replicate. **(b)** Host resistance over time, represented as the proportion of phage genotypes from all time points that were resisted by a host genotype in each replicate. Means and 95% CIs are shown (*N* = 8).

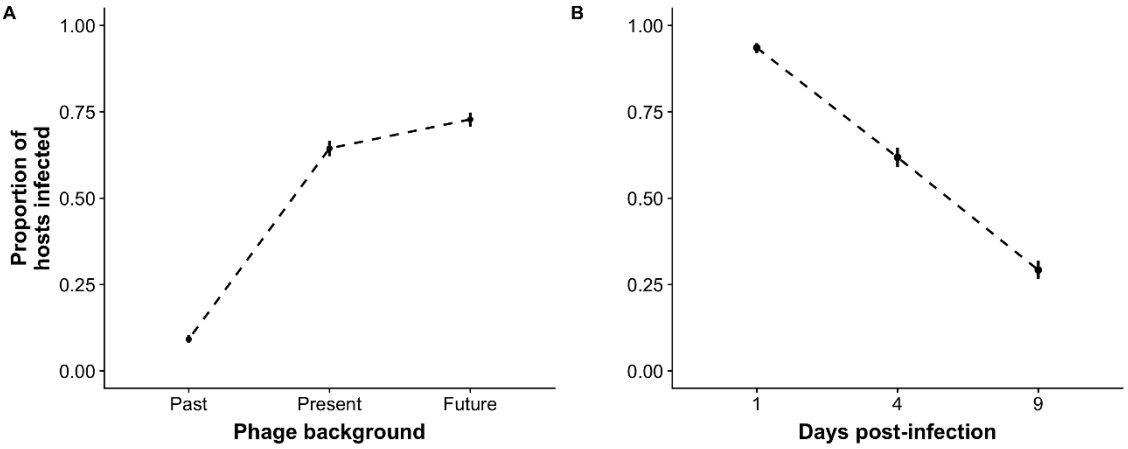


Figure 3. Results from time-shift experiment. **(a)** Proportion of hosts infected when phage were from the host's past, present or future. **(b)** Proportion of hosts infected by phage from the same time point (days post-infection). The dotted line is for illustrative purposes. Means and 95% CIs are shown (*N* = 100048).

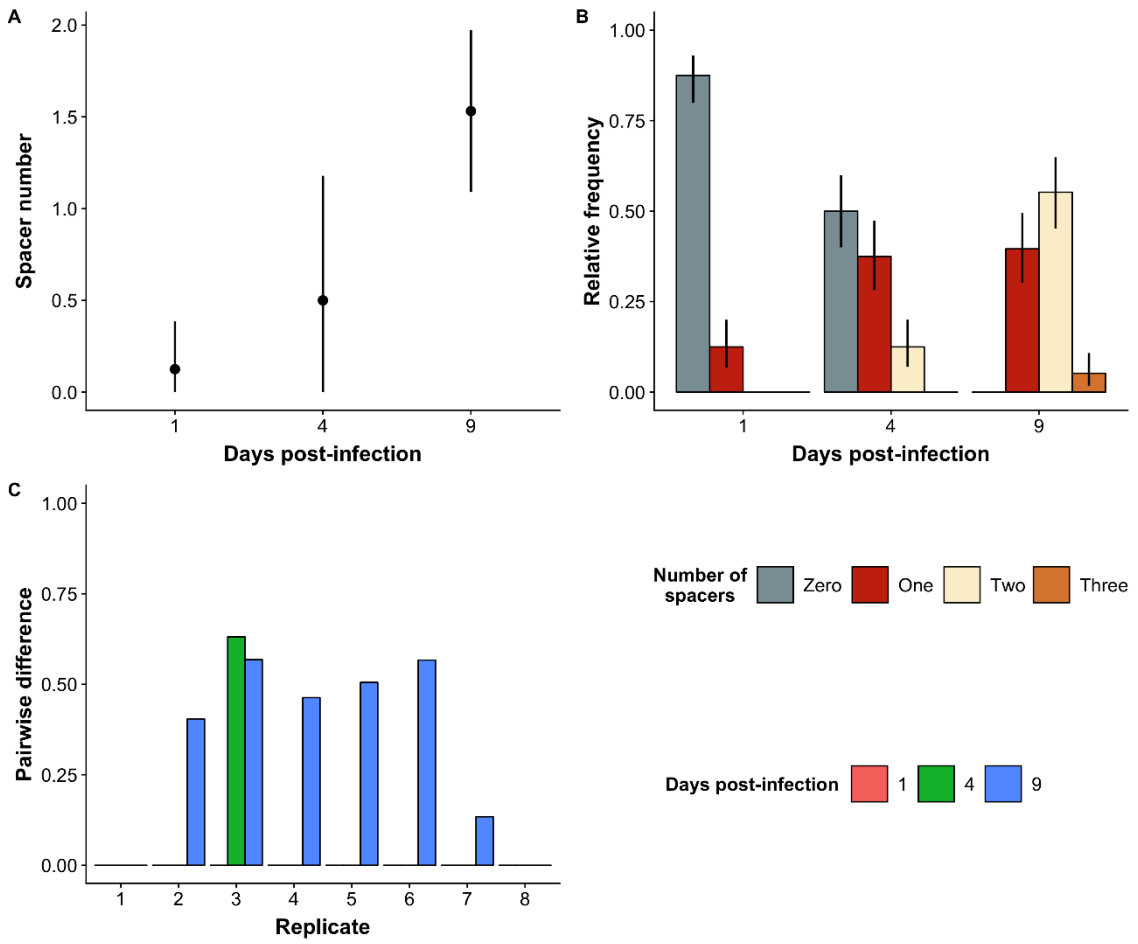


Figure 4. Spacers acquired during coexistence of *S. thermophilus* and phage 2972. **(a)** Number of acquired spacers per clone at each day post-infection (dpi). Means and 95% CIs are shown (*N* = 8). **(b)** Mean relative frequency of clones with different numbers of acquired spacers at each dpi. No clone with >3 spacers was detected. Means and 95% CIs are shown (*N* = 8). **(c)** Spacer diversity in each replicate, measured as the pairwise differences among spacer sequences in each replicate (x-axis) at each time point (colours).

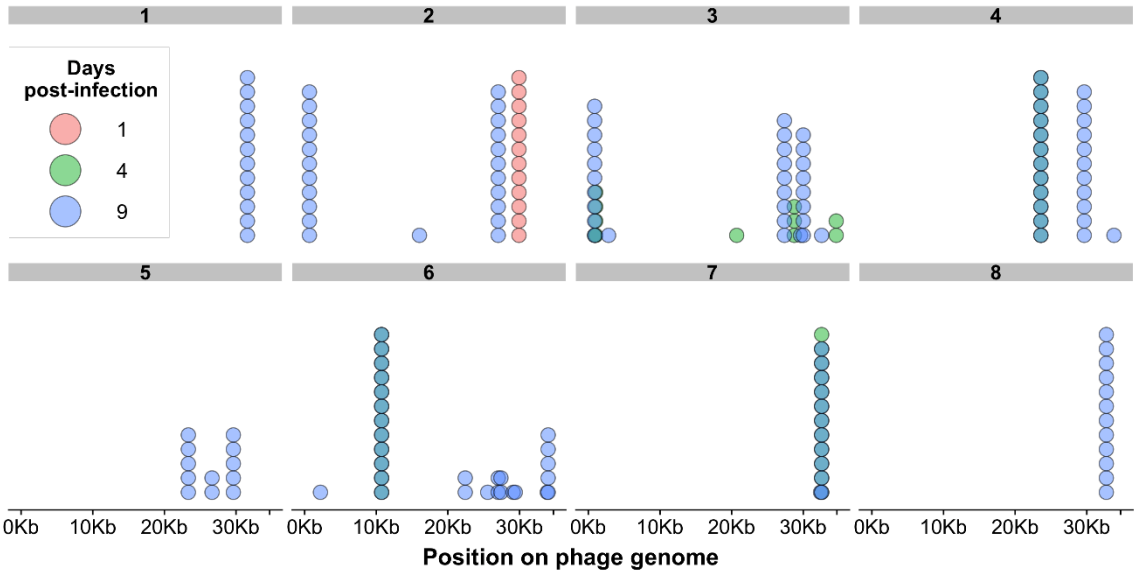


Figure 5. Protospacer locations of newly-acquired spacers. Histogram showing the location of acquired spacers in each replicate when mapped against the phage 2972 genome. Each dot represents a clone that had a spacer mapped to that region. Red, green and blue indicate 1, 4, and 9 days post-infection, respectively. Darker colours are the result of visual overlap between dots. Replicate identity is indicated above each sub-panel.

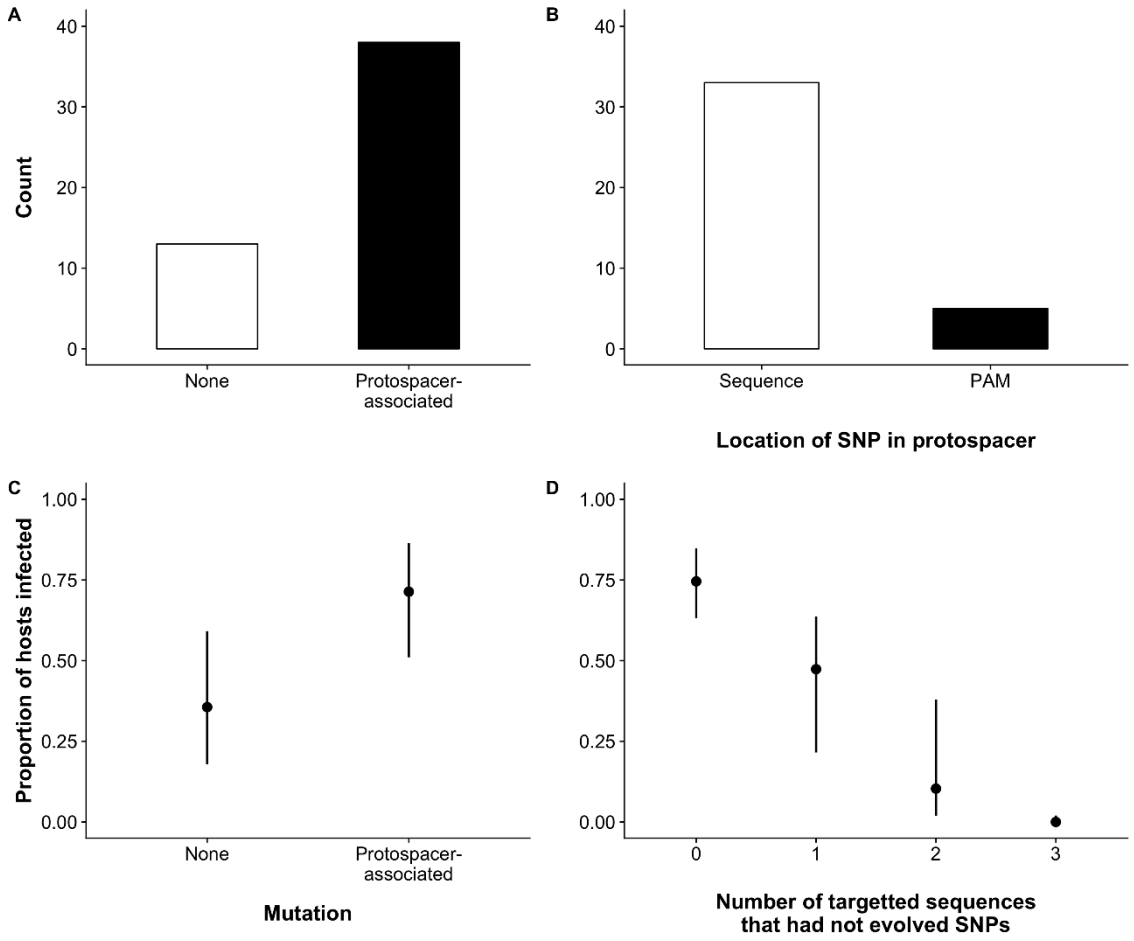


Figure 6. Protospacer sequence analysis and infectivity patterns. **(a)** Histogram showing the number of sequenced phage (out of 56) that did not have a detectable mutation, had a 'random' mutation outside of the protospacer, or had a 'protospacer-associated' mutation either in the protospacer-adjacent motif (PAM) or the seed sequence. **(b)** Number of sequenced phage with protospacer-associated mutations (out of 40) that had a single nucleotide polymorphism (SNP) in either the seed sequence or PAM. **(c)** Mean proportion of hosts infected by phage that did or did not have a protospacer-associated SNP. The effect of random mutations is not included due to limited sample size. **(d)** Mean proportion of hosts infected by phage that had a full or partial match to the host's CRISPR array in terms of the number of targeted protospacer sequences that had not evolved by point mutation. Means and 95% CIs are shown (*N* = 696).

Tables

		Phage		
		T1	**T4**	**T9**
Host	**T1**	Present	Future	Future
	T4	Past	Present	Future
	T9	Past	Past	Present

Table 1. Pairwise challenges between hosts and phage in the time-shift assay. Numbers indicate the time points (days post-infection) analysed. Each row represents whether hosts from that time point were contemporaneous or not with respect to the phage they were challenged with.

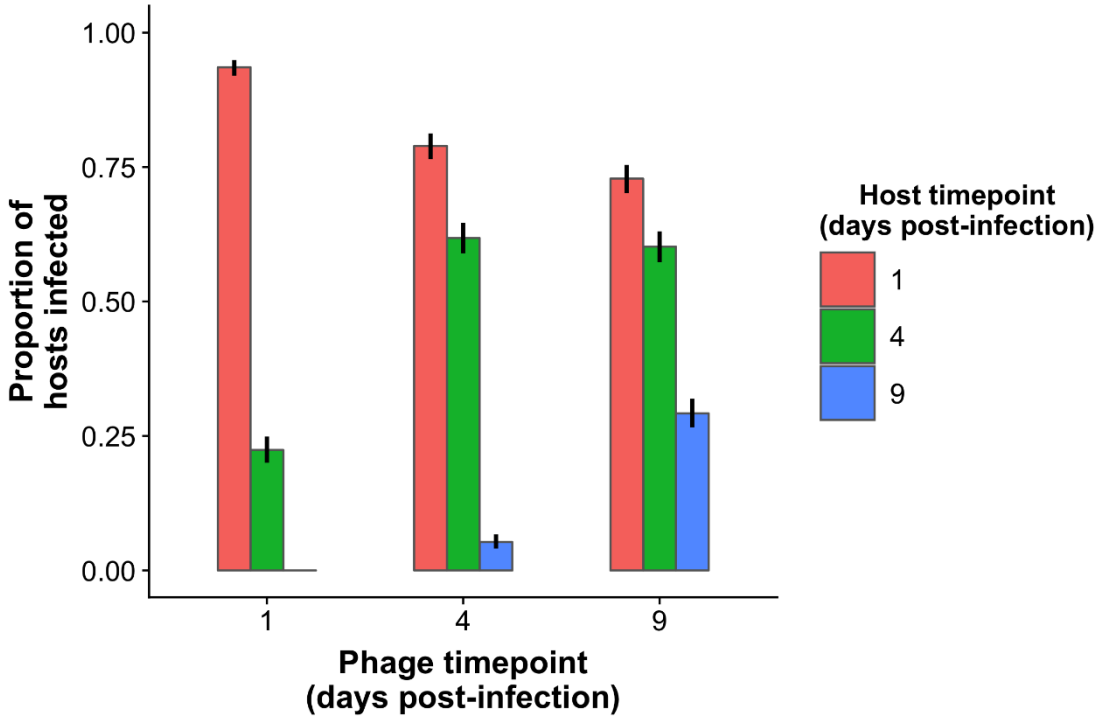
Host origin (day)	Phage origin (day)	Mean Infectivity	Infectivity 95% CI
1	1	0.93	0.92 - 0.95
4	4	0.22	0.20 - 0.25
9	9	0	0.00 - 0.00
1	1	0.79	0.76 - 0.81
4	4	0.62	0.59 - 0.65
9	9	0.05	0.04 - 0.07
1	1	0.73	0.70 - 0.75
4	4	0.6	0.57 - 0.63
9	9	0.29	0.27 - 0.32

Table 2. Mean proportion and 95% confidence interval (CI) of hosts infected and phage resisted in pairwise challenges in the time-shift experiment, broken down by the day from which the host or phage originated. Values are rounded to two decimal places.

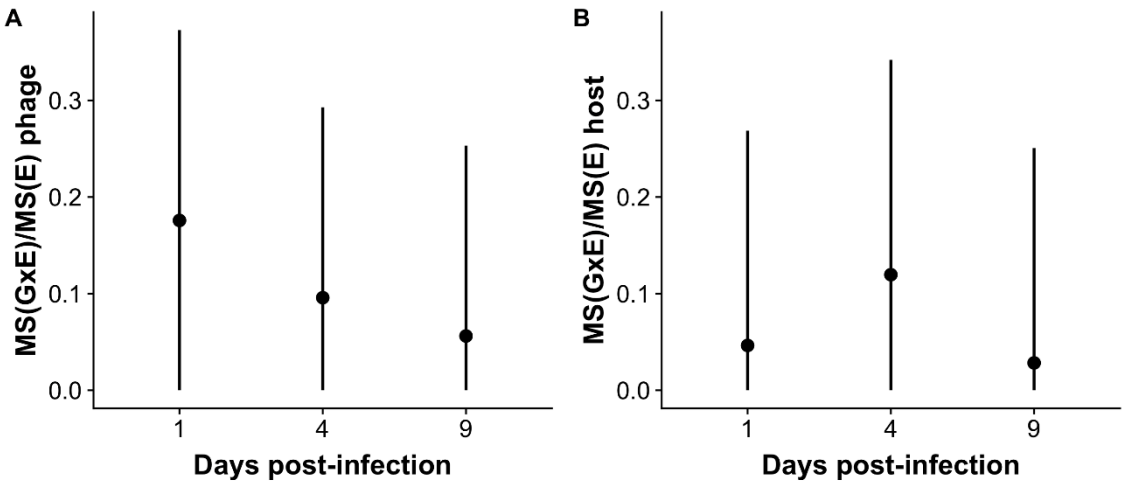
		Predicted	
		-	+
Measured	-	242	107
	+	106	241

Table 3. Contingency table of pairwise infections that were predicted to lead to phage escape based on protospacer sequence data, compared against the pairwise infections that were measured from the phenotypic assay. + indicates a successful infection, - indicates no infection.

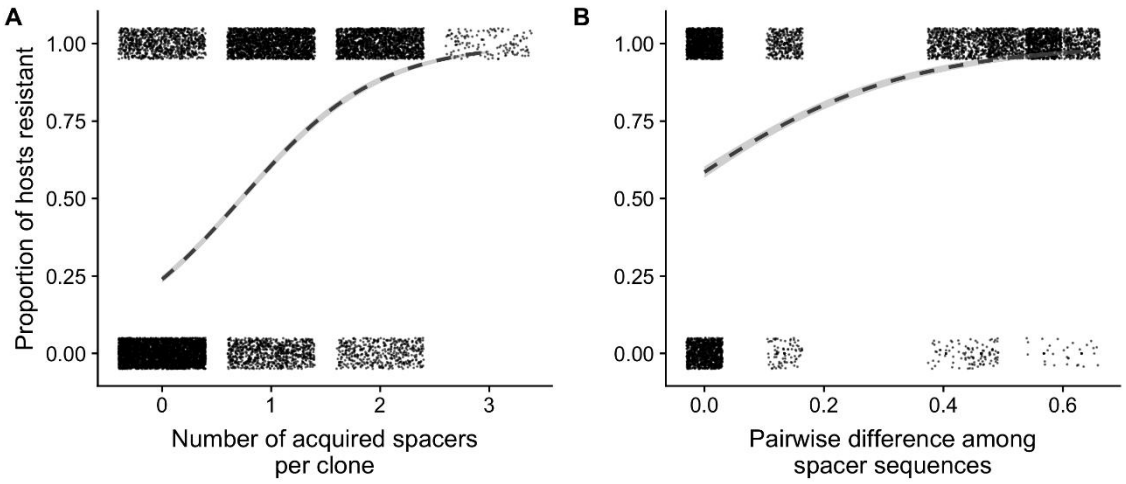
Supplementary Figures



Supplementary Figure 1. Proportion of hosts resistant to phage that were from the host's past, present or future from the time-shift experiment. Means and 95% CIs are shown (*N* = 10048).



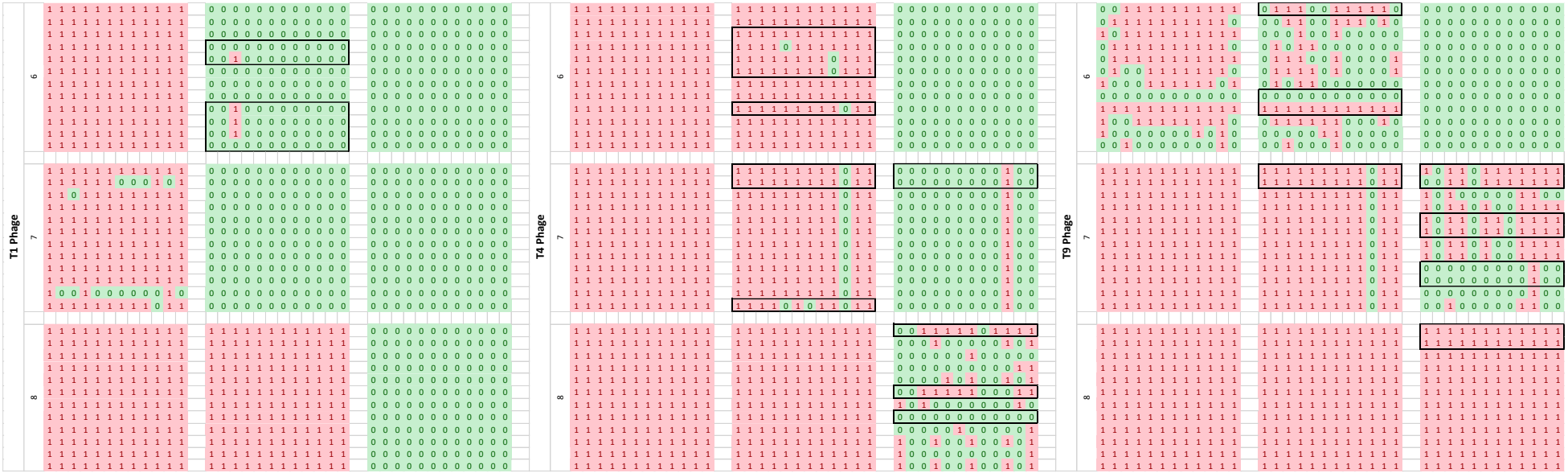
Supplementary Figure 2. Importance of fluctuating selection dynamics (FSD) relative to arms race dynamics (ARD); higher values indicate stronger FSD (see Hall *et al*. 2011, Figure 3 for a more detailed description of this test). Under FSD, pathogen genotypes should differ in their infectivity to hosts from contemporary and non-contemporary environments. Values represent the ratio of the variance in infectivity due to host environment alone explained by variance in infectivity among genotypes, derived from GLMMs with environment as a fixed effect and phage genotype as a random effect. Residuals were square-root transformed to introduce a normal distribution in line with model assumptions. Means and 95% CIs are shown are shown (*N* = 8).



Supplementary Figure 3. Relationship of host susceptibility to **(a)** the number of spacers acquired per clone and **(b)** sequence diversity in terms of the pairwise difference among spacer sequences from the phenotypic data. Points show raw data per clone, with random noise added to give a better indication of the number of occurrences of a given measurement (random noise is represented by dot-filled blocks; darker blocks indicate more occurrences). The dashed line is the smoothed logistic (binomial) regression slope fitted through the data. 95% confidence intervals are show in grey (*N* = 10048).

Supplementary Tables

(continued overleaf)



Supplementary Table 1. Infectivity matrices from the phenotypic assay (see *Methods*). The timepoint from which hosts and phage originated is shown. Each 12x12 block is a time-shift challenge within each replicate (1-8). In each block, columns 1-12 are individual bacterial clones and rows 1-12 are individual phage isolates. 1 = plaque visible, measured infective; 0 = no plaque visible, measured resistant. Phage clones selected for protospacer sequence analysis are indicated by black borders.

T	R	L	S	E	Sequence	N
9	1	CR1	31582	31611	CTCAGTCGTTACTGGTGAACCAGTTTCAAT	12
9	2	CR1	647	676	ATATCGTCCAGACTATCGCAGAATACTGAT	11
9	2	CR1	16002	16032	TATGGTGTCGAAACTTTTAAACGTTACAAT	1
9	2	CR1	27003	27032	AGATTTATAACATGGAAATTGACGATGAAA	11
1	2	CR1	29896	29923	TTATGGAGATGGTTGATTACGCAATCAACT	12
4	3	CR1	867	900	GTTTTAAGTGGTATTATTATATTATCGAAG	1
9	3	CR1	900	867	AGCTTACAATGCGGCTCTTAAAGCTGGATA	10
4	3	CR1	1020	1049	CATAGCTTACAATGCGGCTCTTAAAGCTGG	4
9	3	CR1	2846	2811	TTTTTCGATGAGATATATCAGTACCGATGG	1
4	3	CR1	20696	20725	TGGAATAAGGTATCGTGCGTTTTGACAAGC	1
9	3	CR1	27356	27388	AACACTCAAAGAGTTACTTAAATCTGGAAAG	9
4	3	CR1	28734	28767	AGCTATAGTATATACACATAGCGTAGAAGC	3
9	3	CR1	29618	29647	TTTTCCGTCTTCTTTTTTAGCAAAGATACG	1
9	3	CR1	30017	29988	GCGACTGTTTGGTGGTTACTGACTTTTGCT	8
9	3	CR1	32550	32584	GTATCAAACCAACGTCCATCAGCCATTC	1
4	3	CR1	34587	34616	AGCCTAGATAGCGAAGTTGATCGTATCTAT	2
4	4	CR1	23571	23600	TTGTTAAAAGAAGCACTAGAGGTGATTTAC	12
9	4	CR1	23600	23571	CATTTAGTGGAGATCACGAAGAAAATTGTT	12
9	4	CR1	29618	29647	TTTTCCGTCTTCTTTTTTAGCAAAGATACG	11
9	4	CR1	33769	33799	CTGCGTGGAACTGTCAGAACATAGTAGACTG	1
9	5	CR1	23363	23334	ATAGTTAACGCCTTTACACCGATCGAGGAA	5
9	5	CR1	26669	26639	GCATACCACAGGTATGACCAAAAACAAGAAA	2
9	5	CR1	29618	29647	TTTTCCGTCTTCTTTTTTAGCAAAGATACG	5
9	5	CR1	2214	2185	GGTCATCACCATTAACTAATCGAATAAGAT	1
4	6	CR1	10717	10748	ACGAAGGCTTGGAAAACATTTGATGGTAAC	12
9	6	CR1	10717	10748	ACGAAGGCTTGGAAAACATTTGATGGTAAC	12
9	6	CR1	22423	22395	ACATGATCTACAACTAGGTCAAGATTGCT	2
9	6	CR1	25561	25531	AAACAAAAAATCTTTGAAGTTTATGACATAC	1
9	6	CR1	27003	27032	TTTCATCGTCAATTTCCATGTTATAAATCT	2
9	6	CR1	27410	27377	CTGGAAAGCATATTGAGGGAGCTACTCTTG	2
9	6	CR1	29124	29095	CAAAATAATTGTGGAAAATCACTGGTAAGT	1
9	6	CR1	29407	29378	AACAACCATTATTTGGGTTGGCCCGAATAT	1
9	6	CR1	33877	33848	TGGTAACTGAAAGGTCAGTGGAACGGCACG	1
9	6	CR1	33998	33968	GTACAGAATTATTGAGGAGTTTATTGAACCT	5
9	7	CR3	32408	32436	TCAGAATGGCTGATGGACGTTGGTTTGATAC	1
9	7	CR3	32550	32584	TCAGAATGGCTGATGGACGTTGGTTTGATAC	11
4	7	CR3	32550	32584	TCAGAATGGCTGATGGACGTTGGTTTGATAC	12

9	8	CR1	32707	32736	TGAAAAAACGAGGAGCACTCGTAGGAGTGG	12

Supplementary Table 2. Unique host CRISPR spacers detected by PCR analysis. Replicate (R) and time point (T) (days post-infection) of the clone(s) which had a given spacer are shown, along with the locus (L) of the spacer either (CRISPR1 (CR1) or CRISPR3 (CR3)) and the start (S) and end (E) locations on the phage 2972 genome to which the sequences mapped, given in base pairs. *N* is the number of clones in a replicate X time point combination (max. 12) that had acquired that spacer.

T	R	Phage ID	SNP in protospacer	SNP location 1	SNP location 2	SNP location 3
1	2	5	1	29896		
1	2	6	1	29899		
4	2	1	1	29899		
4	2	2	0			
4	2	3	0			
4	2	4	0			
4	2	5	0			
4	3	1	0			
4	3	2	1	1039		
4	3	3	0			
4	3	4	1	1039		
4	3	10	1	1027		
4	3	11	0			
4	6	3	1	10724		
4	6	4	1	10724		
4	6	5	1	10724		
4	6	6	1	10724		
4	6	9	1	10724		
4	7	1	1	32580		
4	7	2	1	32583		
4	7	12	1	32580		
4	8	1	0			
4	8	6	0			

4	8	8	0			
9	1	7	1	31615		
9	1	8	1	31615		
9	2	1	1	682		
9	2	5	1	682		
9	3	2	1	893	1014	
9	3	3	0			
9	3	9	1	893	1023	29983
9	3	10	0			
9	3	11	1	893	1024	1027
9	3	12	1	893	1023	
9	4	2	0			
9	4	3	1	29614	29626	
9	4	4	1	29614		
9	4	6	1	29614		
9	4	8	1	29614	29623	
9	6	1	1	10724		
9	6	8	1	10724		
9	6	9	1	10724		
9	7	1	1	32587		
9	7	2	1	32587		
9	7	5	1	32587		
9	7	6	1	32583		
9	7	9	1	32587		
9	7	10	0			
9	8	1	1	32747		
9	8	2	1	32747		
0	n/a	1	0			
0	n/a	2	0			
0	n/a	3	0			
0	n/a	4	0			
0	n/a	5	0			
0	n/a	6	0			

0	n/a	7	0
0	n/a	8	0
0	n/a	9	0
0	n/a	10	0
0	n/a	11	0
0	n/a	12	0
0	n/a	13	0
0	n/a	14	0
0	n/a	15	0
0	n/a	16	0
0	n/a	17	0
0	n/a	18	0

Supplementary Table 3. Locations of SNPs detected by PCR when mapped to the phage 2972 genome. Replicate (R) and time point (T) in days post-infection of the phage isolates are shown (time point 0 refers to ancestral phage). Phage ID is the number of the phage isolate (rows 1-12 in the infectivity matrices in Supplementary Table 1). SNP in protospacer indicates if the SNP(s) was in the seed sequence or protospacer-adjacent motif of the target protospacer. Locations are given in base pairs.

Chapter 4

General Discussion

This thesis has explored the antagonistic interactions between bacteria with CRISPR-Cas immunity (*P. aeruginosa* and *S. thermophilus*) and their lytic phages (DMS3vir and phage 2972 respectively) and found that both systems varied significantly in their coevolutionary dynamics. In *P. aeruginosa*, phages were driven extinct by CRISPR within 4-5 days of exposure due to the diversity of CRISPR spacers generated in the host population. However, when spacer diversity was low, phages could become locally adapted (Morley *et al.* 2017). In contrast, phages did coevolve in populations of *S. thermophilus* for up to 30 days following an arms race dynamic caused by increasing CRISPR spacer acquisition and subsequent phage point mutations in the regions targeted by the spacers. These findings are consistent with previous work (Paez-Espino *et al.* 2013; Paez-Espino *et al.* 2015) and provide compelling evidence that CRISPR spacer diversity plays a pivotal role in determining the outcome of CRISPR-phage interactions.

CRISPR's propensity to generate diversity is thought to be a key factor in its evolution as an immune response, as each successive new spacer added to the host population increases overall mean resistance and decreases overall mean phage infectivity (Morley *et al.* 2017; van Houte *et al.* 2016b). It stands to reason that CRISPR systems that can generate high levels of spacer diversity should be better able to prevent phage infections and less likely to result in coevolution than CRISPR systems where diversity-generating potential is low. In natural environments, CRISPR spacer diversity is restricted by the length of the protospacer adjacent motif (PAM); a short sequence of nucleotides on the phage genome that guides Cas proteins to protospacer regions during CRISPR adaptation. Although the PAM sequence is highly-conserved, it can vary in length between CRISPR systems which affects the number of matching sequences (and thus the number of potential protospacers) available to target on the phage's genome. In other words, longer PAM requirements result in fewer potential protospacers and lower overall spacer diversity.

In our experimental lines, the Type I-F CRISPR system found in *P. aeruginosa* has a PAM length of 3 nt (NGG) (Mojica *et al.* 2009), compared to the Type II-A system in *S. thermophilus* which has a PAM length of 5-7 nt (NGGNG and NNAGAAW) (Deveau *et al.* 2008; Horvath *et al.* 2008). The shorter PAM requirement in *P. aeruginosa* meant these bacteria could generate higher levels of spacer diversity relative to *S. thermophilus*. In addition, the *P. aeruginosa* strain UCBPP-PA14 used in our experiments contained a pre-existing CRISPR spacer that partially matched a protospacer on the phage DMS3vir genome. This phenomenon, referred to as 'priming', leads to an accelerated uptake of spacers from the phage (Datsenko *et al.* 2012; Fineran & Charpentier 2012; Swarts *et al.* 2012) and may help to explain the rapid phage extinctions we observed in this system. In contrast, *S. thermophilus* is not thought to be primed against phage 2972. Our experiments reflect these differences in spacer uptake: the total number of CRISPR spacers increased over time in both model systems, but the rate of spacer acquisition was far more pronounced in *P. aeruginosa* where clones evolved five or more spacers within seven days of infection (See Figure 2 in Westra *et al.* (2015)) compared to *S. thermophilus* where a maximum of three spacers evolved over a nine-day period (Chapter 3, Figure 4b).

However, if shorter PAM sequences lead to increased host diversity, and increased host diversity helps to reduce phage infectivity and prevent epidemics; this begs the question: why don't all CRISPR-Cas systems have short PAM requirements? Aside from protospacer recognition, PAM sequences are also important for autoimmunity because they are not present on the host's CRISPR array (Deveau *et al*. 2008). Therefore, PAM length is likely a trade-off between efficient targeting of phage protospacers and reducing the likelihood of self-cleavage. Furthermore, CRISPR systems that are too efficient can prevent the uptake of potentially beneficial DNA (Almendros *et al*. 2012).

Taken together, it is possible to explain the different coevolutionary outcomes we observed in both systems based on the spacer diversity generated by CRISPR. Initial exposure to phage triggers a CRISPR response in bacteria that leads to the uptake of CRISPR spacers in the host population. At this point there are three potential outcomes: 1) CRISPR is able to generate sufficient diversity in the host population to drive the phage extinct (Chapter 2, Figure 1a), 2) Phage completely

overwhelm and destroy the bacteria (Chapter 3, Figure 1a), or 3) CRISPR is able to generate enough diversity to reduce mean phage infectivity but not completely remove them from the population (Chapter 3, Figure 1b-d). In the latter scenario, phage may continue to evolve point mutations to 'escape' any novel CRISPR spacers emerging in the host population, but their mutations are limited by increasing fitness costs (Lenski & Levin 1985), whereas the fitness cost imposed on bacteria to acquire another CRISPR spacer is negligible (Vale *et al.* 2015). In addition, bacteria require only a partial match from a single spacer to prevent phage from infecting, whereas phage require mutations in all the protospacer sequences targeted by the spacers to remain infectious. Under such circumstances it is plausible that bacteria and phage become entangled in a coevolutionary arms race that permits their coexistence over long periods and can lead to local adaptation, as we observed in our experiments with *S. thermophilus*.

These findings could be of use to industries that rely on the large-scale growth or metabolic activities of bacteria; for example, in the food industry cheese and yoghurt are produced by the fermentation of milk in large vats by specific strains of lactic acid-producing bacteria (including *S. thermophilus*) that are prone to phage contamination. It is common for cheesemakers to create starter cultures of bacterial strains with increased phage resistance to mitigate the risks of phage contamination (Marcó *et al*. 2012). The results of our experiments suggest that choosing bacteria with a diverse range of CRISPR spacers would be beneficial to any starter cultures in this instance as it would reduce the likelihood of a phage epidemic spreading through the population. These results could also be useful in a medical context; particularly in phage-based therapies where phages are used to combat antibiotic-resistant infections in patients where traditional antibiotics have failed or are not effective. One of the major challenges in phage therapy is trying to select phage(s) that are cost-effective and have few side effects whilst maintaining high infection specificity in the face of evolving bacterial resistance. A common approach is to create cocktails containing different phage strains and species to maximise their effectiveness, but these cocktails can be expensive to produce and can impact on normal 'healthy' microbiota leading to undesirable side effects. An understanding of how phages coevolve with CRISPR can help to

select phages with a higher chance of clearing the bacteria before they can evolve resistance.

Collectively, these experiments have provided an insight into the complex antagonistic interactions between bacteria and phage and explained how CRISPR can lead to coevolution when spacer diversity is low. It would be interesting to see how the coevolutionary dynamics change over even longer time periods to see if the arms race we observed in *S. thermophilus* gives way to fluctuating selection from a lack of spacer diversity or due to increasing phage mutation costs, or whether it is inevitable that the CRISPR will eventually prove too difficult for the phage to overcome. Future coevolution experiments with other CRISPR bacteria such as *E. coli* would help with our understanding of CRISPR-phage coevolution and determine whether these results can be generalised beyond the model systems used in our studies.

Bibliography

Agrawal A, Lively CM (2002) Infection genetics: gene-for-gene versus matching alleles models and all points in between. *Evolutionary Ecology Research* **4**, 79–90.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. *Czaki, Akademiai Kiado, Budapest*.

Almendros C, Guzmán NM, Díez-Villaseñor C, *et al*. (2012) Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in Escherichia coli. *PLoS One* **26**, 7(11):e50797.

Amitai G, Sorek R (2016) CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Micro* **14**, 67-76.

Andersson AF, Banfield JF (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047-1050.

Ashby B, King KC (2015) Diversity and the maintenance of sex by parasites. *J. Evol. Biol.* **28**, 511-520.

Barrangou R, Fremaux C, Deveau H*, et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712.

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Berg Miller ME, Yeoman CJ, Chia N*, et al.* (2012) Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol* **14**, 207-227.

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429.

Buckling A, Brockhurst M (2012) Bacteria–virus coevolution. In: *Evolutionary systems biology*, pp. 347-370. Springer.

Buckling A, Rainey PB (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society of London B: Biological Sciences* **269**, 931-936.

Burnham KP, Anderson DR (2003) *Model selection and multimodel inference: a practical information-theoretic approach* Springer Science & Business Media.

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research* **33**, 261-304.

Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA (2012) The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *Journal of bacteriology* **194**, 5728-5738.

Carte J, Christopher RT, Smith JT*, et al.* (2014) The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in S treptococcus thermophilus. *Molecular microbiology* **93**, 98-112.

Chabas H, van Houte S, Høyland-Kroghsbo NM, Buckling A, Westra ER (2016) Immigration of susceptible hosts triggers the evolution of alternative parasite defence strategies. *Proc. R. Soc. B* **283**, 20160721.

Chaudhry WN, Pleška M, Shah NN*, et al.* (2018) Leaky resistance and the conditions for the existence of lytic bacteriophage. *PLoS biology* **16**, e2005971.

Childs LM, England WE, Young MJ, Weitz JS, Whitaker RJ (2014) CRISPR-induced distributed immunity in microbial populations. *PLoS One* **9**, e101710.

Childs LM, Held NL, Young MJ, Whitaker RJ, Weitz JS (2012) Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. *Evolution: International Journal of Organic Evolution* **66**, 2015-2029.

Datsenko KA, Pougach K, Tikhonov A*, et al.* (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications* **3**, 945.
Deveau H, Barrangou R, Garneau JE*, et al.* (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology* **190**, 1390-1400.
Emerson JB, Andrade K, Thomas BC*, et al.* (2013) Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013**, 370871.
Fineran PC, Blower TR, Foulds IJ*, et al.* (2009) The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proceedings of the National Academy of Sciences* **106**, 894-899.
Fineran PC, Charpentier E (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202-209.
Gaba S, Ebert D (2009) Time-shift experiments as a tool to study antagonistic coevolution. *Trends in Ecology & Evolution* **24**, 226-232.
Gandon S, Buckling A, Decaestecker E, Day T (2008) Host–parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology* **21**, 1861-1866.
Gandon S, Nuismer SL (2009) Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *Am Nat* **173**, 212-224.
Gómez P, Ashby B, Buckling A (2015) Population mixing promotes arms race host–parasite coevolution. *Proceedings of the Royal Society of London B: Biological Sciences* **282**.
Gorter FA, Scanlan PD, Buckling A (2016) Adaptation to abiotic conditions drives local adaptation in bacteria and viruses coevolving in heterogeneous environments. *Biology Letters* **12**.
Greischar MA, Koskella B (2007) A synthesis of experimental work on parasite local adaptation. *Ecol Lett* **10**, 418-434.
Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics* **8**, 172.
Hall AR, Scanlan PD, Morgan AD, Buckling A (2011) Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecology Letters* **14**, 635-642.
Held NL, Whitaker RJ (2009) Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. *Environ Microbiol* **11**, 457-466.
Horvath P, Coûté-Monvoisin A-C, Romero DA*, et al.* (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *International journal of food microbiology* **131**, 62-70.
Horvath P, Romero DA, Coûté-Monvoisin A-C*, et al.* (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of bacteriology* **190**, 1401-1412.
Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biometrical journal* **50**, 346-363.
Iranzo J, Lobkovsky AE, Wolf YI, Koonin EV (2013) Evolutionary dynamics of archaeal and bacterial adaptive immunity systems, CRISPR-Cas, in an explicit ecological context. *Journal of bacteriology*, JB. 00412-00413.
Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology* **169**, 5429-5433.
Jansen R, Embden JDv, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**, 1565-1575.
Janzen DH (1966) Coevolution of Mutualism between ants and acacias in Central America. *Evolution*, **20,** 249-275.
Jiang W, Maniv I, Arain F*, et al.* (2013) Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS genetics* **9**, e1003844.
Kaltz O, Shykoff JA (1998) Local adaptation in host-parasite systems. *Heredity* **81**, 361-370.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.

Kearse M, Moir R, Wilson A*, et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.
Keesing F, Belden LK, Daszak P*, et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647.
King KC, Lively CM (2012) Does genetic diversity limit disease spread in natural host populations? *Heredity* **109**(4), 199.
Koonin EV, Makarova KS, Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* **37**, 67-78.
Koskella B (2014) Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution. *The American Naturalist* **184**, S9-S21.
Koskella B, Lively CM (2007) Advice of the rose: experimental coevolution of a trematode parasite and its snail host. *Evolution* **61**, 152-159.
Kunin V, He S, Warnecke F*, et al.* (2008) A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18**, 293-297.
Laanto E, Hoikkala V, Ravantti J, Sundberg L-R (2017) Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nature communications* **8**, 111.
Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8**, 317.
Laine AL (2008) Temperature-mediated patterns of local adaptation in a natural plant-pathogen metapopulation. *Ecol Lett* **11**, 327-337.
Lenski RE, Levin BR (1985) Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *The American Naturalist* **125**, 585-602.
Levin BR, Moineau S, Bushman M, Barrangou R (2013) The population and evolutionary dynamics of phage and bacteria with CRISPR–mediated immunity. *PLoS genetics* **9**, e1003312.
Lively CM (2010) An epidemiological model of host–parasite coevolution and sex. *Journal of Evolutionary Biology* **23**, 1490-1497.
Lively CM (1999) Migration, Virulence, and the Geographic Mosaic of Adaptation by Parasites. *The American Naturalist* **153**, S34-S47.
Lopez Pascua L, Gandon S, Buckling A (2012) Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. *Journal of evolutionary biology* **25**, 187-195.
Marcó MB, Moineau S, Quiberoni A. (2012) Bacteriophages and dairy fermentations. *Bacteriophage*. **2**(3), 149-58.
Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics* **11**, 181.
Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740.
Mojica FJ, García-Martínez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174-182.
Morgan AD, Gandon S, Buckling A (2005) The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* **437**, 253-256.
Morley D, Broniewski JM, Westra ER, Buckling A, van Houte S (2017) Host diversity limits the evolution of parasite local adaptation. *Molecular ecology* **26**, 1756-1763.
Orsini L, Schwenk K, De Meester L*, et al.* (2013) The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution* **28**, 274-282.

Paez-Espino D, Morovic W, Sun CL*, et al.* (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature communications* **4**, 1430.
Paez-Espino D, Sharon I, Morovic W*, et al.* (2015) CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. *MBio* **6**, e00262-00215.
Plagens A, Richter H, Charpentier E, Randau L (2015) DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiology Reviews* **39**, 442-463.
R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Råberg L, Alacid E, Garces E, Figueroa R (2014) The potential for arms race and Red Queen coevolution in a protist host–parasite system. *Ecology and evolution* **4**, 4775-4785.
Sasaki A (2000) Host-parasite coevolution in a multilocus gene-for-gene system. *Proc Biol Sci* **267**, 2183-2188.
Semenova E, Jore MM, Datsenko KA*, et al.* (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences* **108**, 10098-10103.
Sorek R, Kunin V, Hugenholtz P (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology* **6**, 181.
Sorokin VA, Gelfand MS, Artamonova, II (2010) Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. *Appl Environ Microbiol* **76**, 2136-2144.
Sun CL, Thomas BC, Barrangou R, Banfield JF (2016) Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *The ISME journal* **10**, 858.
Swarts DC, Mosterd C, Van Passel MW, Brouns SJ (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888.
Therneau TM, Lumley T (2015) Package 'survival'. *R Top Doc* **128**.
Thompson JN (1994) *The coevolutionary process* University of Chicago Press.
Thompson JN (1999) The evolution of species interactions. *Science* **284**, 2116-2118.
Thompson JN (2005) *The geographic mosaic of coevolution* University of Chicago Press.
Tollrian R, Harvell D, Harvell CD (1999) *The ecology and evolution of inducible defenses* Princeton University Press.
Tyson GW, Banfield JF (2008) Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10**, 200-207.
Vale PF, Lafforgue G, Gatchitch F*, et al.* (2015) Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. *Proc. R. Soc. B* **282**, 20151270.
Vale PF, Little TJ (2010) CRISPR-mediated phage resistance and the ghost of coevolution past. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 2097-2103.
Van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in biochemical sciences* **34**, 401-407.
van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Micro* **12**, 479-492.
van Houte S, Buckling A, Westra ER (2016a) Evolutionary ecology of prokaryotic immune mechanisms. *Microbiology and Molecular Biology Reviews* **80**, 745-763.
van Houte S, Ekroth AK, Broniewski JM*, et al.* (2016b) The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385-388.
Weinberger AD, Wolf YI, Lobkovsky AE, Gilmore MS, Koonin EV (2012) Viral diversity threshold for adaptive immunity in prokaryotes. *MBio* **3**, e00456-00412.
Weissman JL, Holmes R, Barrangou R*, et al.* (2018) Immune loss as a driver of coexistence during host-phage coevolution. *The ISME journal* **12**, 585.
Westra ER, Dowling AJ, Broniewski JM, van Houte S (2016) Evolution and ecology of CRISPR. *Annual Review of Ecology, Evolution, and Systematics* **47**, 307-331.

Westra ER, Semenova E, Datsenko KA*, et al.* (2013) Type IE CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS genetics* **9**, e1003742.
Westra ER, Sünderhauf D, Landsberger M, Buckling A (2017) Mechanisms and consequences of diversity-generating immune strategies. *Nature Reviews Immunology* **17**, 719.
Westra ER, Swarts DC, Staals RH*, et al.* (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* **46**, 311-339.
Westra ER, van Houte S, Oyesiku-Blakemore S*, et al.* (2015) Parasite exposure drives selective evolution of constitutive versus inducible defense. *Current Biology* **25**, 1043-1049.
Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York. Version.
Wilson GG, Murray NE (1991) Restriction and modification systems. *Annual Review of Genetics* **25**, 585-627.